

## Article

# A Dataset Quality Assessment—An Insight and Discussion on Selected Elements of Environmental Footprints Methodology

Anna Lewandowska <sup>1</sup>, Katarzyna Joachimiak-Lechman <sup>1,\*</sup>  and Przemysław Kurczewski <sup>2</sup>

<sup>1</sup> Department of Quality Management, Poznań University of Economics and Business, 61-875 Poznań, Poland; anna.lewandowska@ue.poznan.pl

<sup>2</sup> Faculty of Civil and Transport Engineering, Poznań University of Technology, Piotrowo 3, 60-965 Poznań, Poland; przemyslaw.kurczewski@put.poznan.pl

\* Correspondence: katarzyna.joachimiak-lechman@ue.poznan.pl

**Abstract:** One of the most recently developed life cycle-based methods is an environmental footprint of products and organisations established by the European Commission. A special procedure of data and dataset quality assessment has been developed as a part of the environmental footprints methodology. The procedure may be recognised as vital and powerful but, at the same time, a bit complicated and time-consuming. It is worth discussing this subject and looking for potential simplification. In this paper, we suggest a possible way for simplification. We propose to remove an impact-assessment-based step from the procedure of company-specific datasets quality assessment. There are two potential benefits: a reduction in the need for expert knowledge and time savings. The threats posed are connected to the fact that all data influences the Data Quality Rating indicator of the entire dataset to the same degree. With a higher volume of data included in the assessment, there is a risk of greater differentiation in their quality. In this paper, an example of raw milk production is presented. The assessment of quality of the dataset was performed in three variants: pursuant to the approach established by the European Commission in the pilot phase, transition phase and with certain modifications employed.

**Keywords:** Product Environmental Footprint; company-specific datasets; data quality assessment



**Citation:** Lewandowska, A.; Joachimiak-Lechman, K.; Kurczewski, P. A Dataset Quality Assessment—An Insight and Discussion on Selected Elements of Environmental Footprints Methodology. *Energies* **2021**, *14*, 5004. <https://doi.org/10.3390/en14165004>

Academic Editor: Marek Dudek

Received: 9 June 2021

Accepted: 6 August 2021

Published: 15 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Many environmental problems are related to energy. It is extremely important to assess the consequence of energy production and consumption in the whole life cycle perspective. Life cycle-based environmental tools, such as the Environmental Life Cycle Assessment (LCA), Product Carbon Footprint (CFP), Product Water Footprint (WFP), or Product Environmental Footprint (PEF), are becoming increasingly important in both policy strategies and economic practice [1–4]. Life cycle-based management tools are also widely applied in the energy sector [5–7]. The reliability of the final results obtained in life cycle analyses is influenced by many factors from different phases of these studies, e.g., assumptions and value choices made by the practitioner when defining the goal and scope, multifunctional solutions in product systems, data quality (LCI and LCIA), model quality (LCI and LCIA), weighting factors, study reviews, or even the version and update of secondary databases. Despite the credibility of final study results based on the life cycle concept constituting a component for many variables, one of its key elements is the quality of data used in phase two: the analysis of the input and output data sets (LCI), i.e., inventory data. Because energy occurs in many company-specific datasets and often seems to be a hotspot in the product's life cycle, it is important to analyse the procedures established to assess the quality of company-specific datasets.

### 1.1. Data Quality Assessment—General Overview

There are procedural steps related to the quality of inventory data to be taken at three of the four phases of the life cycle analysis: definition of the goal and scope (data quality requirements), a life cycle inventory (validation of data: data quality assessment and treatment of missing data) and an interpretation (uncertainty analysis). Data quality is understood to be ‘characteristics of data that relate to its ability to satisfy the stated requirements’. Thus, the starting point is to define these requirements. ISO standards concerning life cycle techniques provide for the parameters that should be included in the data quality requirements. Moreover, over the years, these parameters have not changed a great deal. The first version of ISO 14040:1997 listed seven parameters, while the most recent ISO 14044:2006 provides for ten data quality parameters: time-related coverage; geographical coverage; technology coverage; precision; completeness; representativeness; consistency; reproducibility; sources of the data; and the uncertainty of the information [8]. In the case of comparative assertions intended to be disclosed to the public, all of the parameters mentioned above should be addressed. With a strong dependence of life cycle analyses on study context and the need for flexibility, the requirements and guidelines in ISO standards are still universal, but general. Additionally, with regards to the assessment of data quality, the standards do not provide detailed instructions. Therefore, over the years, some approaches have been developed in which these general guidelines have been operational. The first publications on data quality assessment appeared in the early 1990s, i.e., even before the first ISO 14040s standards were published. In 1992, SETAC published a conceptual framework for Life-Cycle Assessment Data Quality [9]. In 1995, the EPA commissioned a report entitled Guidelines for assessing the quality of Life Cycle Inventory Analysis [10]. Vigon and Jensen published the results of their survey of individuals and organisations experienced in data and database quality assessment [11] and, a year later, Weidema and Wesnaes’ work on an example of using data quality indicators was published [12]. These publications provide a systematic summary of data sources, data types, data aggregation, data quality goals, and data quality indicators. Weidema and Wesnaes proposed using the so-called Pedigree Matrix as a semi-quantitative approach to data quality assessment with five data quality indicators and a five-point scale for scoring. Weidema published the results of testing the proposed approach in 1998 [13].

Until recently, the Pedigree Matrix had been successfully applied in practice. Many publications [14–20] refer to the concept of data quality indicators and the Pedigree Matrix. Van den Berg et al. (1999) provided an example of the operationalisation of the framework for quality assessment in LCA, based on the Spread-Assessment Pedigree, by evaluating the overall quality of an LCA result with the use of 15 different quality factors related to unit processes or whole systems [14]. The ILCD Handbook includes practical aspects of the data quality concept, quality levels and quality ratings for the data quality indicators, among others [15]. Papers [16,17] present the methodological issues and results of an eco-invent project to refine the Pedigree Matrix approach. Some methodological considerations concerning existing methods used to assess the quality of the LCA study are also presented in Lewandowska et al. (2004) [18]. The updated data quality system included an approach to the Pedigree Matrix, as described in Guidance on Data Quality Assessment for Life Cycle Inventory Data [19]. Additional recommendations on data quality creation, management and use in LCA databases and studies are included in Edelen and Ingwersen (2018) [20]. One of the latest proposals based on this semi-quantitative approach includes a data quality assessment methodology, developed by the Joint Research Centre for the European Commission as part of the Environmental Footprints Initiative for products and organisations [21].

### 1.2. Environmental Footprints—Basic Aspects

According to the Commission Recommendation (2013/179/EU), the motivation for the development of the environmental footprints methodology was to address the problem of proliferation of different methods and initiatives used to assess and communicate life cy-

cle environmental performance. Environmental footprints have a common methodological core for Product Environmental Footprint (PEF) [22,23] and Organisation Environmental Footprint (OEF) [22,24], while more operational and detailed guidelines have been developed for individual product categories (Product Environmental Footprint Category Rules, PEFCRs) and sectors (Organisation Environmental Footprint Sector Rules, OEFSRs).

Environmental footprints are based on the LCA methodology and draw some inspiration from the requirements and guidelines of ISO 14040s. The difference is that while the ISO guidelines are basically general, the guidelines for environmental footprints are planned to be more detailed and to constitute a kind of ‘cookbook’ for PEF/OEF practitioners. This also applies to data and datasets’ quality assessment. A first procedure was developed by the European Commission during the Environmental Footprint (EF) pilot phase (2013–2018) [25] while the updated version has been published during the current transition phase [23,24]. The EF pilot phase was aimed at testing the processes for creating product and sector-specific rules, testing various approaches to verification and testing different communication vehicles [21]. The EF transition phase has been focused on three goals: ‘monitoring the implementation of existing PEFCRs/OEFSRs; developing new PEFCRs/ OEFSRs; and new methodological developments’ [21].

### 1.3. Aim of the Study

In this paper, we will apply both versions of the procedure to assess the quality of a company-specific dataset developed by the European Commission: (1) during the pilot phase and (2) during the transition phase. This procedure seems to be a powerful approach but, at the same time, it may be recognised as being too complex and time-consuming. It is worth discussing this subject and looking for potential simplifications. Our contribution is to suggest a possible way for simplification. We propose to remove an impact-assessment-based step of identification of the most relevant issues from the procedure. There are two potential benefits from the simplification: a less expert knowledge-based procedure (easier to be made by non-LCA experts e.g., suppliers sharing primary data) and time savings.

An example of company-specific data is presented for the annual production of a dairy farm, modelled in situation 2, option 1, pursuant to the Data Needs Matrix [23]. The presented example is hypothetical and intended to illustrate data quality-related considerations, not to assess the environmental performance of the farm. The dataset includes 18 direct elementary flows (dEFs) and 16 activity data (AD). The activity data refer to consumption of feed components (maize, barley, straw), water, electricity, heat and waste management. Different geographical locations of supplier activity were assumed. The quality of the entire dataset was assessed by using three approaches: pursuant to the approach established by the European Commission in the EF pilot phase, in the EF transition phase and with a suggested minor modification employed. The calculated quality has been expressed as a Data Quality Rating (DQR) value and compared with the minimum level allowed for a company-specific dataset [23].

Two main limitations of our case study need to be highlighted. Due to some restrictions in free-to-use, EF-compliant secondary datasets, all activity data were modelled with datasets taken from the ecoinvent 3.6 database. These datasets are not EF compliant but this does not interfere with the presentation or explanation of the concept of our proposal; however, it must clearly be stated that it is a kind of noncompliance. EF-compliant, secondary datasets shall be applied in real PEF calculations. Additionally, as a consequence of using non-EF compliant secondary datasets, a modified version of the EF impact assessment method was used in the case study (adapted EF method 2.0). Some of the results are included in the Supplementary Information, which is an inherent element of this paper.

## 2. Quality of Company-Specific Data(Set)—A Summary of Guidelines

### 2.1. Company-Specific Data—A Quality Assessment

In EF studies, a quality assessment procedure separately refers to the data itself and the related datasets. Both are strictly related but not the same. A PEF practitioner assesses

the quality of self-collected inventory data. Following the analysis of guidelines in [23], it is possible to differentiate the following cases, in which (depending on the situation) company-specific data shall be or may be collected:

- with reference to processes, activity data and direct elementary flows which, in PEFCR, have been included in the list of mandatory company-specific data (this pertains to products, for which PEFCR exists);
- with reference to processes that are run by the company performing PEF;
- with reference to processes, that are not run by the company performing PEF, but
- the company has access to primary data.

The situations are discussed in more detail in Table 1. As can be seen, it is of high relevance whether the analysis is performed for a product belonging to a product category covered by PEFC or not. PEFCR provides two basic pieces of information: a list of mandatory company-specific data and a list of the most relevant issues (including the most relevant processes and elementary flows). Both matters determine the data requirements. Further criteria (common for products covered and not covered by PEFCR) are supervision and operational control over the process and access to specific data. A detailed description of the quality requirements for processes in the PEF study is included in the Data Needs Matrix (DNM), indicated in Zampori's and Pant's report [23]. In this report, the DNM variant may be found for products covered and not covered by PEFCR.

The PEFCR (e.g., Annex 6) classifies inventory information into three categories: (1) mandatory company-specific, (2) expected to be company-run and (3) secondary. If a PEF study is intended to be disclosed as compliant with the PEFCR, then company-specific data shall be collected for all mandatory company-specific items. This list of mandatory company-specific data included in a PEFCR (Annex 6) should be considered mandatory.

The 'expected to be company run' category is context-dependent and includes processes that may or may not be under the direct control of the company performing PEF and with or without access to primary data. As such, the minimum level of quality of 'expected to be company run' data depends on the context of the study. PEFCRs also indicate some processes for which using secondary information is allowed. Whenever a PEF practitioner decides to gather company-specific data for a process not included in the list of mandatory company-specific data, then the data quality assessment procedure presented below shall also be used. It is the same procedure for all company-specific data, regardless of whether they are collected for a mandatory company-specific process or for others.

The quality of company-specific data is expressed as a Data Quality Rating (DQR) value and shall be assessed by using several criteria. The lower the DQR value, the better the data quality. Four criteria have been included [23,25] for PEFCR: precision (P)—a measure of the variability of the data values for each of the data expressed; time representativeness (TiR)—age of data and the minimum length of time over which data should be collected; technological representativeness (TeR)—characterises to what degree the used data is depicting the technology of the system analysed; and geographical representativeness (GR)—characterises to what degree the used data is depicting the geography of the system analysed. The guidelines presented in Table 2 should be used for activity data (AD) and elementary flows (EFs). The total DQR for each individual data point is calculated as the weighted average of each criterion ( $DQR = \frac{P+TiR+TeR+GR}{4}$ ). DQR = 1 corresponds to

excellent quality and is achieved if a rating of 1 is obtained for all four criteria. DQR = 2.5 is the lowest possible score for company-specific data and is to be achieved in case of having the following scores for particular criteria: 3.0 for precision, 3.0 for time representativeness, 2.0 for technological representativeness, and 2.0 for geographical representativeness.

According to the PEF guidance documents, it is acceptable for the Technical Secretariat to include time representativeness-adapted data in the PEFCR guidelines (TiR-EF and TiR-AD) [23,25]. However, this is the only criterion that may differ in a given PEFCR; the others must be as shown in Table 2 [23,25]. This table is used to assess company-specific data quality for both PEFCR and non-PEFCR products.

**Table 1.** Eight modelling situations with differentiated minimum needs for collecting company-specific data.

Is the Product under Study Covered by Any Existing PEF CR?	Is the Process on the List of Mandatory Company-Specific Data?	Is the Process Run by the Company Performing PEF?	Does the Company Performing PEF Have Access to Primary Data?	Is the Process on the List of the Most Relevant Processes (Indicated in the PEF CR)?	Instruction for PEF Practitioner on Company-Specific Data Collection
Yes	Yes	n.a.	n.a.	n.a.	For this process, company-specific data (related to activity data and direct elementary flows) shall be collected and assessed. No other option available.
	No	Yes	Yes	Yes	For this process, company-specific data (related to activity data and direct elementary flows) shall be collected and assessed. No other option available.
	No	Yes	Yes	No	For this process, as one of two options, company-specific data (related to activity data and direct elementary flows) may be collected and assessed, Instead, using secondary data(set) is possible.
	No	No	Yes	Yes/No	For this process, company-specific data (related to activity data and direct elementary flows) should be collected and assessed. Two variants of data collection are available: (1) full primary inventory or (2) using secondary data(set) with primary data related only to electricity and transport.
	No	No	No	No	Using secondary data(set) is recommended.
No	n.a.	Yes	Yes	n.a.	For this process, company-specific data (related to activity data and direct elementary flows) shall be collected and assessed. No other option available.
	n.a.	No	Yes	n.a.	For this process, company-specific data (related to activity data and direct elementary flows) should be collected and assessed. Two variants of data collection are available: (1) full primary inventory or (2) using secondary data(set) with primary data related only to electricity and transport.
	n.a.	No	No	n.a.	Using secondary data(set) is recommended.

Source: Elaboration based on [23].

**Table 2.** Data Quality Rating for company-specific information—according to Zampori and Pant [23].

	Rating	Precision	Time Representativeness	Technological Representativeness	Geographical Representativeness
		$P_{EF}$ and $P_{AD}$	$Ti_{R-EF}$ and $Ti_{R-AD}$	$Te_{R-EF}$ and $Te_{R-AD}$	$G_{R-EF}$ and $G_{R-AD}$
the best quality	1	The activity data and elementary flows are based on measurements/calculations, and externally verified	The activity data and elementary flows apply for most recent annual administration period (in relation to the EF report publication date)	The elementary flows and the activity data refer exactly to the technology of the newly developed dataset	The activity data and elementary flows reflect the exact geographical location of the process is modelled in the newly created dataset
	2	The activity data and elementary flows are based on measurements/calculations, and internally verified, plausibility checked by a reviewer	The activity data and elementary flows apply for the maximum two annual administration periods (in relation to the EF report publication date)	The elementary flows and the activity data are a proxy of the technology of the newly developed dataset	The activity data and elementary flows partly reflect the geographical location of the process is modelled in the newly created dataset
	3	The activity data and elementary flows are based on measurements/calculations/taken from literature and plausibility not checked by a reviewer OR Qualified estimate based on calculations' plausibility checked by a reviewer	The activity data and elementary flows apply for the maximum three annual administration periods (in relation to the EF report publication date)	Not applicable	Not applicable
the worst quality	4–5	Not applicable	Not applicable	Not applicable	Not applicable

Source: Elaboration based on [23].

## 2.2. Company-Specific Datasets (CSD)—A Quality Assessment

A mandatory task for PEF practitioners is not only to gather company-specific data and assess its quality, but to also use this data to create company-specific datasets and assess the quality of the datasets. The criteria presented in Table 2 indicate that the worst allowed quality rating for a single company-specific data item is  $DQR = 2.5$ . As such, the following question is raised: ‘Is it possible to use all company-specific data with the worst  $DQR = 2.5$ ?’. The answer is ‘no, it is impossible’. The reason is that the PEF quality assessment procedure includes minimum quality levels required for the entire company-specific datasets. According to the procedure developed by the European Commission during the EF transition phase [23], the minimum level of quality of an entire company-specific dataset has to be equal to or lower than 1.5 ( $DQR_{\text{Company\_specific\_dataset\_transition}} \leq 1.5$ ). For comparison, pursuant to the methodology elaborated earlier in the pilot phase [25], the threshold was 1.6 ( $DQR_{\text{Company\_specific\_dataset\_pilot}} \leq 1.6$ ). In practice, this means that the quality of company-specific data may be differentiated (ranged between  $DQR = 1$  and  $DQR = 2.5$ ); however, ultimately, the total DQR for the whole dataset shall be  $DQR \leq 1.5$ .

Following Zampori & Pant’s report [23], the procedure of the CSD quality assessment established during the transition phase (for products with and without PEFCR) presents as follows:

- Step 1. Calculate the environmental impact of the dataset (weighted results, toxicity impact categories included and absolute values). Identify the most relevant AD and dEFs: the most relevant activity data are the ones linked to sub-processes (i.e., secondary datasets) that account for at least 80% of the total environmental impact of the company-specific dataset, listing them from the most contributing to the least contributing. Most relevant direct elementary flows are defined as those direct elementary flows cumulatively contributing at least with 80% to the total impact of the direct elementary flows [23].
- Step 2. Calculate the DQR criteria  $Te_R$ ,  $Ti_R$ ,  $Ge_R$  and  $P$  for each most relevant activity data and each most relevant direct elementary flow. The values of each criterion shall be assigned based on the table on how to assess the value of the DQR criteria provided in the PEFCR or in the PEF method (in our paper, presented in Table 2).
- Step 3. Calculate the environmental contribution of each most-relevant activity data (through linking to the appropriate sub-process) and direct elementary flow in the total score, to be calculated as a sum of the environmental impact of all most-relevant activity data and direct elementary flows, in % (weighted, using all EF impact categories).
- Step 4. Calculate the  $Te_R$ ,  $Ti_R$ ,  $Ge_R$  and  $P$  criteria of the newly developed dataset as the weighted average of each criterion of the most relevant activity data and direct elementary flows. The weight is the relative contribution (in %) of each most relevant activity data and direct elementary flow calculated in step 3.
- Step 5. Calculate the total DQR of the newly created dataset:

$$DQR = \frac{\overline{Ti_R} + \overline{Te_R} + \overline{Ge_R} + \overline{P}}{4} \quad (1)$$

where  $\overline{Ti_R}$ ,  $\overline{Te_R}$ ,  $\overline{Ge_R}$  and  $\overline{P}$  are the weighted averages, calculated as specified in step 4.

For comparison, an analogical procedure established during the pilot phase was as follows [25]:

- Step 1. Calculate the environmental impact of the dataset (weighted results, toxicity impact categories excluded and absolute values). Identify the most relevant AD and dEFs: most relevant activity data are the ones linked to sub-processes (i.e., secondary datasets) that account for at least 80% of the total environmental impact of the company-specific dataset, listing them from the most contributing to the least contributing. Most relevant direct elementary flows are defined as those direct elementary flows contributing, cumulatively, at least 80% to the total impact of the direct elementary flows.

- Step 2. Calculate the DQR criteria  $Te_R$ ,  $Ti_R$ ,  $G_R$  and  $P$  for each most relevant process and each most relevant direct elementary flow. The values of each criterion shall be assigned based on the requirements presented in Table 2. For each most relevant elementary flow, evaluate the DQR for four criteria:  $Te_{R-EF}$ ,  $Ti_{R-EF}$ ,  $G_{R-EF}$  and  $P_{EF}$ . The quality of each most relevant process is a combination of the quality of activity data and the quality of the secondary dataset used.  $Ti_{R-AD}$  and  $P_{AD}$  shall be evaluated at the level of the activity data and  $Te_{R-SD}$ ,  $Ti_{R-SD}$  and  $G_{R-SD}$  shall be assessed at the level of the secondary dataset used. As  $Ti_R$  is evaluated twice, the mathematical average of  $Ti_{R-AD}$  and  $Ti_{R-SD}$  shall be calculated.
- Step 3. Calculate the environmental contribution of each most-relevant activity data (through linking to the appropriate sub-process) and direct elementary flow to the total sum of the environmental impact of all most-relevant activity data and direct elementary flows, in % (weighted, without toxic impact categories).
- Step 4. Calculate the  $Te_R$ ,  $Ti_R$ ,  $G_R$  and  $P$  criteria of the newly developed dataset as the weighted average of each criterion of the most relevant activity data and direct elementary flows. The weight is the relative contribution (in %) of each most relevant activity data and direct elementary flow calculated in step 3.
- Step 5. Calculate the total DQR of the newly created dataset:

$$DQR = \frac{\overline{Ti_R} + \overline{Te_R} + \overline{Ge_R} + \overline{P}}{4} \quad (2)$$

where  $\overline{Ti_R}$ ,  $\overline{Te_R}$ ,  $\overline{Ge_R}$ , and  $\overline{P}$  are the weighted averages, calculated as specified in step 4.

The principal similarities between the two procedures are:

- the quality rating of individual AD and EFs in a dataset is performed with the use of the same criteria (Table 2);
- the calculation of the final DQR indicator for a dataset is indispensably connected with LCIA calculations and the obtained LCIA results;
- according to the procedure for the identification of the most relevant issues presented in Zampori & Pant's report (Section 6.3.5) [23] and PEFCR Guidance (Section 7.4.5) [25], indicating the most relevant AD and EFs is carried out by the use of absolute results for environmental indicators, neglecting negative values;
- the value of the final DQR indicator is 'powered by' the quality of inventory elements with the highest environmental relevance (a materiality approach).

The principal differences between the two procedures are:

- updated guidelines [23] include instructions on how to make the quality assessment both for products with existing PEFCR and without existing PEFCR. The PEFCR Guidance [25] presents guidelines for PEF studies to be performed with PEFCR only;
- the minimum allowed DQR for company-specific datasets and mandatory company-specific data is  $\leq 1.5$  in the updated guidelines [23] and  $\leq 1.6$  in the PEFCR Guidance [25];
- in the updated guidelines [23], only data quality (the most relevant AD and dEFs) is taken into account during the calculation of total DQR for the entire company-specific dataset. In PEFCR Guidance [25], the quality of the company-specific dataset is calculated as a combination of the quality of data (the most relevant AD and dEFs) and the quality of secondary datasets referred to the most relevant AD;
- in order to assess the quality of company-specific datasets, the identification of the most relevant activity data and the most relevant elementary flows is to be performed without toxic impact categories in the case of the pilot procedure, while the transition procedure includes the toxic impact categories.

### 3. Proposal for Simplification of the Company-Specific Datasets Quality Rating

We propose the modification of the company-specific dataset quality assessment procedure through the introduction of the three following steps:

- Step 1. Calculate the DQR criteria  $Te_R$ ,  $Ti_R$ ,  $Ge_R$  and  $P$  for each activity data and direct elementary flow included in the dataset. The values of each criterion shall be assigned

based on the table on how to assess the value of the DQR criteria provided by the PEFCR or by PEF method (see Table 2).

- Step 2. Calculate the TeR, TiR, GeR and P criteria of the newly developed dataset as the arithmetic mean of each criterion of all activity data and all direct elementary flows.
- Step 3. Calculate the total DQR of the newly created dataset:

$$\text{DQR} = \frac{\overline{Ti_R} + \overline{Te_R} + \overline{Ge_R} + \overline{P}}{4} \quad (3)$$

where  $\overline{Ti_R}$ ,  $\overline{Te_R}$ ,  $\overline{Ge_R}$  and  $\overline{P}$  are the averages, as calculated in step 2.

In step 1, it has been assumed that the same primary data rating criteria are used as in the approaches discussed earlier (Table 2), but the quality rating of the entire dataset was separated from LCIA calculations. This procedure calls for a quality rating of all inputs and outputs in a set, without consideration of their environmental relevance. Thus, the share of each inventory item in the final DQR for a dataset is equally relevant.

#### 4. Three Approaches in Use—An Example of the Quality Assessment of a Company-Specific Dataset for Raw Milk Production

Let us assume that the PEF study is executed for yoghurt by a producer, in the reference year 2019, the PEF report publication date is in June 2020 and the study was commissioned in 2019. The study concerns the supplier of raw milk, who agreed to deliver primary data. The production of raw milk is not run by the dairy unit but the facility has access to primary data shared by the farmstead. Yoghurt falls into the category of ‘fermented milk products’ and is covered by a valid PEFCR for dairy products [26]. Pursuant to this document, the production of raw milk is the most relevant process. This means that, according to the Data Needs Matrix, this process should be modelled in situation 2, option 1, and the minimum allowable quality level for this dataset is  $\text{DQR} \leq 1.5$  (according to the pilot phase procedure, it is  $\leq 1.6$ ). Annex 6 of the PEFCR for dairy products provides the data requirements and includes a list of activity data and elementary flows to be collected for dairy farms. All of the data has been classified as ‘Expected to be company-run (only for companies with direct access to dairy farmers such as cooperatives)’ [26].

In our example, we assess the quality of data obtained from a single farmstead using the criteria in Table 2 for P, TeR and GR. The raw milk suppliers’ sampling selection (sampling) is not considered and the quality of the data obtained from a larger number of farmsteads is not considered. It must be noted, however, that PEFCR [26] contains guidelines for the assessment of primary data obtained from a sample of farmsteads. They cover three criteria (TiR, TeR and GR). TeR and GR are closely interrelated with the size and structure of a sample and, in terms of temporal representativeness, the criteria have been defined in the manner presented in Table 3 [26].

As may be noted, the criteria for TiR in Table 2 refer to the EF report publication date and, conversely in Table 3, to the year in which the EF study was commissioned. Additionally, within the scope of TiR in Table 2, only the age of the data is considered, without reference to the possible averaging over several-year periods. Criteria presented in Table 3 seem to be softer, as the highest quality indicator is possible to be obtained for data with a deviation of 5 years towards the time of commissioning the study. On the other hand, the universal criteria for TiR contained in Table 2 represent the most recent annual administration period with respect to the EF report publication date.

Agricultural production is subject to seasonal fluctuation; therefore, accounting for data covering periods of several years is justified. Due to the fact that, following the PEF method update [23], criteria pertaining to the quality assessment of company-specific data should be based on the guidelines presented in Table 2 and ‘only the reference years criteria (TiR-EF, TiR-AD) may be adapted by the Technical Secretariat.’ In our example for TiR, as a base scenario, the guidelines in Table 3 were used, with the assumption that, in the case of modelling raw milk production in PEF studies (consistently with the PEFCR and within the scope of temporal representativeness (TiR)), they were prioritised over the guidelines in Table 2.

**Table 3.** Time Representativeness—Data Quality Rating for company-specific information for raw milk production according to the PEFCR for dairy products [26].

Rating	Time Representativeness	
	Ti <sub>R-EF</sub> and Ti <sub>R-AD</sub>	
the best quality	1	The average calculated based on the production data covering over 2+ years period, not older than 5 years, in relation to the year the study was commissioned
	2	The average calculated based on the production data covering over 2+ years period, not older than 10 years, in relation to the year the study was commissioned
	3	The average calculated based on the production data for a single year, in the previous 5 years, in relation to the year the study was commissioned
	4	The average calculated based on the production data for a single year, in the previous 10 years, in relation to the year the study was commissioned
the worst quality	5	Production data for an unknown period or a period lower than 1 year

Source: Elaboration based on [26].

We assume that the analysed supplier delivered 168,867 kg of FPCM milk. In the Supplementary Information file in Supplementary Table S1, the supplier's assumed characteristics (for the purpose of our analysis) are presented. Supplementary Table S2 includes an example of a company-specific dataset for averaged operation data for the farmstead in 2018 and 2019. For the purpose of simplification and omission of allocation matters, we assumed that the analysed farmstead performs only animal rearing (without growing crops) and does not sell meat or manure.

In this example, we assume that all pieces of information pertaining to inputs, as well as waste and sewage, were measured or collected from the farmstead's documentation and are averaged values from two operational years (2018 and 2019). Direct emissions in the farmstead, stemming from intestinal fermentation in animals and the management of manure, were calculated with breeding parameters, emission indicators and methodology from the following reports: the IPCC Guidelines for National Greenhouse Gas Inventories [27,28], EMEP/EEA emission inventory guidebook [29,30] and a report of the National Centre for Emission Balancing and Management [31,32]. We assume that all data were verified internally and checked by a reviewer. In Supplementary Table S2, we have separately provided elementary flows, even if they cover emissions of the same substance to the same environmental surroundings, as they pertain to various emission sources and are calculated with the use of different indicators, which in practice may result in varying data quality. As shown in Supplementary Table S2, 18 direct elementary flows (dEFs) and 16 activity data (AD) are included in the scope of our company-specific dataset. In real PEF studies, the EF-compliant secondary datasets would be used to model the activity data. In our case study, in place of secondary EF-compliant datasets, all activity data have been modelled with datasets taken from the ecoinvent 3.6 database. This is a deviation, as the used ecoinvent datasets do not meet the EF compliance requirements on modelling, meta-data, nomenclature or data quality rating [33]. In order to identify the most relevant activity data and direct elementary flows, life cycle impact assessment calculations have been made by using an adapted EF 2.0 method v. 1.01.

The quality of the entire dataset was assessed by three approaches. The results of this assessment are presented in Table 4. The assessment was made by one person with some experience in data quality assessment in EF studies. Before the assessment, all of the inventory data was input to a template prepared in an Excel file, where special fields for Data Quality Rating (DQR) values for each AD and EF were created. The cells were provided with a comment explaining the criteria and rating, which made the assessment faster and easier.

In our analysis, we assumed that all data from Supplementary Table S2 had been collected from the farmstead's documentation or based on calculations performed with the use of the emission indicators, as well as being verified internally and checked by a reviewer. For this reason, for the precision criterion P, we assigned all inventory items with a value of DQR = 2. Except for waste and sewage management, all data correspond

to the technology employed in the farmstead; thus, the DQR for parameter  $Te_R$  is 1. For waste, we reduced it to 2, as the segregation into waste designated for incineration and disposal was performed with the indices for Poland, from Annex C [34], which are non-specific for concrete technology, but they represent an average scenario for country. In the scope of  $Ti_R$ , all input flows obtained the best indicator of 1, because it was assumed that they come from documentation and measurements from 2018/2019, and the report was assumed to be published in June 2020. Therefore, these data refer to the most recent annual administration period with respect to the EF report publication date. The direct emissions are calculated based on the farmstead parameters (e.g., milk productivity of cows, body weight of milk cattle, time spent in livestock premises, proportion of silage in the feed, manure management system, etc.), which, in this analysis, corresponded to the technology used in the farmstead in 2018/2019. Both aspects (time of publication of a report with emission indicators, as well as the age of data pertaining to the farmstead's operational parameters) were considered during the evaluation of  $Ti_R$  for dEFs pertaining to emissions in a farmstead. We have used criteria from Table 3 in the base scenario for the purpose of quality assessment within the scope of  $Ti_R$ , which provide for a several-year tolerance with respect to the year of commissioning the study. Therefore, in the base approach, all data pertaining to emissions to air were assessed as having been  $Ti_R = 1$ .

For emissions calculated based on IPCC 2019 emission indicators, the indicators of  $Ge_R$  have been reduced to 2. According to the methodology used, the emission indicators and parameters are indicated for regions of the world (e.g., Western Europe, Eastern Europe, North America), and not for individual countries. For the purpose of our analysis, we have assumed 'Eastern Europe' for Poland, which may constitute a certain underestimation and reduced quality with regard to  $Ge_R$ .

In the base scenario, the final quality indicator for the entire dataset, calculated with the modified approach, amounted to  $DQR_{CDS} = 1.38$ . The rating took 17 min to perform. For an approach elaborated in the transition period, the result for the indicator amounted to  $DQR_{CSD} = 1.28$  and the rating took 72 min. The greatest portion of time was consumed for the analysis performed pursuant to the approach elaborated in the pilot period. Additionally, the quality of the secondary datasets used to model the most relevant AD had to be rated (according to the guidelines presented in the PEFCD for dairy products, in Section 5.5 and Table 30). In this case, the value of  $DQR_{CDS} = 1.79$  was obtained and the rating was performed in 91 min. The necessity of linking quality indicators with LCA results clearly influenced the extension of the time needed for the execution of the rating. Additionally, a contribution analysis for the absolute values had to be performed manually in an Excel sheet. The  $DQR_{CDS}$  of 1.79 is too high and does not satisfy the minimum quality requirements. This results from a weak geographical representativeness of the secondary dataset (RoW) used to model the most relevant process: maize grain production in Poland (PL). Additionally, a temporal validity of the used secondary datasets expired at the end of 2019, i.e., one year before the time assumed for publication of the PEF report.

Higher  $DQR_{CDS}$  values were obtained in cases where quality assessment for  $Ti_R$  criterion was performed, with the guidelines provided in Table 2. The results for both approaches are presented in Table 5. The differences stem from the smaller temporal tolerance allowed by the guidelines presented in Table 2. It is particularly evident in the case of the proposed approach ( $DQR_{CSD} = 1.38$  vs.  $DQR_{CSD} = 1.47$ ), in which the quality of all data is taken into account. Quality indicators of  $Ti_R$  for direct emissions were differentiated, depending on the source reports' publication dates (IPCC, EMEP/EAA, National Centre for Emission Balancing and Management), from which the methodology and emission indicators had been collected. In the approaches of the pilot and transition phases, this was of less importance, because only three emissions to air had been considered to be the most relevant issues (as well as relatively low weights). On the other hand, in the proposed approach, as many as 15 emissions were taken into account for the purpose of the quality assessment.

**Table 4.** An assessment of the quality of data and a company-specific dataset (CSD) for raw milk production made using three approaches ( $T_{IR}$  assessed according to the guidelines presented in Table 3).

Flow Name	Flow Location	Flow Type	Approach Established in the EF Pilot Phase							Approach Established in the EF Transition Phase						Possible Simplification							
			Data Quality Calculated as a Weighted DQR Based on the Quality of the Most Relevant AD/EFs and the Quality of Secondary Datasets Used to Model the Most Relevant AD							Data Quality Calculated as a Weighted DQR Based on the Quality of the Most Relevant AD/EFs						Data Quality Calculated as the Average Quality of All AD and EFs Included in the Dataset (Arithmetic Mean)							
			The Most Relevant? (Based on LCIA Calculations, EF Method 2.0 (Adopted), Without Toxic ICs)	Flow Weight	$P_{EF}$ and $P_{AD}$	$T_{IR-EF}$ and $T_{IR-AD}/T_{IR-SD}$	$T_{ER-EF}$ and $T_{ER-SD}$	$G_{R-EF}$ and $G_{R-SD}$	DQR per Flow (Before Weighing)	The Most Relevant? (Based on LCIA Calculations EF Method 2.0 (Adopted), with Toxic ICs)	Flow Weight W	$P_{EF}$ and $P_{AD}$	$T_{IR-EF}$ and $T_{IR-AD}$	$T_{ER-EF}$ and $T_{ER-AD}$	$G_{R-EF}$ and $G_{R-AD}$	DQR per Flow (Before Weighing)	The Most Relevant? (Based on LCIA Calculations)	Flow Weight	$P_{EF}$ and $P_{AD}$	$T_{IR-EF}$ and $T_{IR-AD}$	$T_{ER-EF}$ and $T_{ER-AD}$	$G_{R-EF}$ and $G_{R-AD}$	DQR per Flow
Land occupation, agricultural	PL	dEF	No						No						n.a.	n.a.	2	1	1	1	1.25		
Land transformation from meadow	PL	dEF	No						No						n.a.	n.a.	2	1	1	1	1.25		
Land transformation to agricultural	PL	dEF	no						No						n.a.	n.a.	2	1	1	1	1.25		
Maize grain	PL	AD	yes	0.45	2	1.5	2	3	2.125	Yes	0.48	2	1	1	1.25	n.a.	n.a.	2	1	1	1	1.25	
Maize grain	CH	AD	yes	0.11	2	1.5	2	1	1.625	Yes	0.15	2	1	1	1.25	n.a.	n.a.	2	1	1	1	1.25	
Straw	PL	AD	yes	0.09	2	1.5	2	2	1.875	Yes	0.10	2	1	1	1.25	n.a.	n.a.	2	1	1	1	1.25	
Maize silage	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Maize silage	BR	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Barley grain	DE	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Additives	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Tap water	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Electricity	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Heat, from natural gas	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Diesel	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
HDPE film	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Transport	PL	AD	no						No						n.a.	n.a.	2	1	1	1	1.25		
Biogenic methane (air)—enteric fermentation	PL	dEF	yes	0.12	2	1	1	2	1.5	Yes	0.10	2	1	1	2	1.5	n.a.	n.a.	2	1	1	2	1.5

Table 4. Cont.

Flow Name	Flow Location	Flow Type	Approach Established in the EF Pilot Phase							Approach Established in the EF Transition Phase					Possible Simplification								
			Data Quality Calculated as a Weighted DQR Based on the Quality of the Most Relevant AD/EFs and the Quality of Secondary Datasets Used to Model the Most Relevant AD							Data Quality Calculated as a Weighted DQR Based on the Quality of the Most Relevant AD/EFs					Data Quality Calculated as the Average Quality of All AD and EFs Included in the Dataset (Arithmetic Mean)								
			The Most Relevant? (Based on LCIA Calculations, EF Method 2.0 (Adopted), Without Toxic ICs)	Flow Weight	P <sub>EF</sub> and P <sub>AD</sub>	T <sub>IR-EF</sub> and T <sub>IR-AD</sub> /T <sub>IR-SD</sub>	T <sub>ER-EF</sub> and T <sub>ER-SD</sub>	G <sub>R-EF</sub> and G <sub>R-SD</sub>	DQR per Flow (Before Weighing)	The Most Relevant? (Based on LCIA Calculations EF Method 2.0 (Adopted), with Toxic ICs)	Flow Weight W	P <sub>EF</sub> and P <sub>AD</sub>	T <sub>IR-EF</sub> and T <sub>IR-AD</sub>	T <sub>ER-EF</sub> and T <sub>ER-AD</sub>	G <sub>R-EF</sub> and G <sub>R-AD</sub>	DQR per Flow (Before Weighing)	The Most Relevant? (Based on LCIA Calculations)	Flow Weight	P <sub>EF</sub> and P <sub>AD</sub>	T <sub>IR-EF</sub> and T <sub>IR-AD</sub>	T <sub>ER-EF</sub> and T <sub>ER-AD</sub>	G <sub>R-EF</sub> and G <sub>R-AD</sub>	DQR per Flow
Biogenic methane (air)—manure management	PL	dEF	yes	0.03	2	1	1	2	1.5	Yes	0.03	2	1	1	2	1.5	n.a.	n.a.	2	1	1	2	1.5
Dinitrogen monoxide (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
Dinitrogen monoxide (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
Dinitrogen monoxide (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
Dinitrogen monoxide (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
Nitrates (water)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
Ammonia (air)	PL	dEF	yes	0.18	2	1	1	1	1.25	Yes	0.14	2	1	1	1	1.25	n.a.	n.a.	2	1	1	1	1.25
Nitrogen oxides (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	1	1.25	
Carbon dioxide, fossil (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	1	1.25	
PM 2.5 (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
NMVOG (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
NMVOG (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
NMVOG (air)	PL	dEF	no							No						n.a.	n.a.	2	1	1	2	1.5	
HDPE waste, incineration	PL	AD	no							No						n.a.	n.a.	2	1	2	1	1.5	
HDPE waste, landfilling	PL	AD	no							No						n.a.	n.a.	2	1	2	1	1.5	
Wastewater treatment	PL	AD	no							No						n.a.	n.a.	2	1	2	1	1.5	
			Company-specific dataset (CSD) quality		P <sub>CSD</sub>	T <sub>IR-CSD</sub>	T <sub>ER-CSD</sub>	G <sub>R-CSD</sub>	DQR <sub>CSD</sub>	Company-specific dataset (CSD) quality		P <sub>CSD</sub>	T <sub>IR-CSD</sub>	T <sub>ER-CSD</sub>	G <sub>R-CSD</sub>	DQR <sub>CSD</sub>	Company-specific dataset (CSD) quality		P <sub>CSD</sub>	T <sub>IR-CSD</sub>	T <sub>ER-CSD</sub>	G <sub>R-CSD</sub>	DQR <sub>CSD</sub>
					2.0	1.3	1.7	2.2	1.79			2.0	1.0	1.0	1.1	1.28			2.0	1.0	1.1	1.4	1.38

Source: Own elaboration.

**Table 5.** DQR of the company-specific dataset for raw milk production— $Ti_R$  calculated using criteria presented in Tables 2 and 3.

Approach Established in the EF Pilot Phase						Approach Established in the EF Transition Phase						Possible Simplification					
Company-Specific Dataset (CSD) Quality	$P_{CSD}$	$Ti_{R-CSD}$	$Te_{R-CSD}$	$G_{R-CSD}$	$DQR_{CSD}$	Company-Specific Dataset (CSD) Quality	$P_{CSD}$	$Ti_{R-CSD}$	$Te_{R-CSD}$	$G_{R-CSD}$	$DQR_{CSD}$	Company-Specific Dataset (CSD) Quality	$P_{CSD}$	$Ti_{R-CSD}$	$Te_{R-CSD}$	$G_{R-CSD}$	$DQR_{CSD}$
TiR assessed based on criteria presented in the Table 3	2.0	1.3	1.7	2.2	1.79	TiR assessed based on criteria presented in Table 3	2.0	1.0	1.0	1.1	1.28	TiR assessed based on criteria presented in Table 3	2.0	1.0	1.1	1.4	1.38
TiR assessed based on criteria presented in the Table 2	2.0	1.5	1.7	2.2	1.85	TiR assessed based on criteria presented in Table 2	2.0	1.1	1.7	1.6	1.32	TiR assessed based on criteria presented in Table 2	2.0	1.41	1.1	1.4	1.47

Source: Own elaboration.

## 5. Discussion

The linking of the data quality assessment with the environmental relevance of the activity data and direct elementary flows features a substantive justification. The limitation of this assessment to the most relevant AD and dEFs is the representation of the materiality principle, this being the basis for the concept of environmental footprints. One advantage is the fact that a relatively small number of AD and dEFs influence the quality of a dataset, and thus the efforts of the body collecting the data may be focused on collecting the highest quality data, not only pertaining to a few of the inventory items. However, the procedure for acquiring information, which is the most relevant out of all inventory data, is time-consuming and potentially very difficult for people having no experience in LCA/EF analyses. What is more, the procedure is executed when, one way or another, the majority of inventory data should be collected. In practice, producers may cooperate with dozens or even hundreds of suppliers. If even only a portion of them decide to deliver a complete set of primary data pertaining to their operation, then the quality rating of these data and datasets, based on the necessity to perform LCIA calculations, would complicate and extend the entire analysis considerably. In addition, it would complicate the execution of the verification and the possibilities to reconstruct and verify the results.

Thus, it seems that, from a purely practical point of view, this conditioning of data and dataset quality rating with the impact assessment features some considerable weaknesses. Potentially, it extends the analysis time and requires a performer to possess expert knowledge and skills to handle, e.g., LCA software. The solution we propose is, in fact, very simple, and what is more, following a short training period and preparation of model sheets (templates), the suppliers of primary data would be capable of evaluating and managing the quality of the delivered data on their own. Maybe, from a substantial point of view, the proposed simplification trivialises the quality rating, but from a practical point of view, it accelerates the rating and simplifies it at the level of performing calculations.

The fact that all AD and dEFs influence the DQR indicator of the entire dataset to the same degree poses a certain threat. With a higher number of data included in the assessment, the risk of greater differentiation in their quality arises. Consequently, it may be more challenging to obtain the required level of dataset quality  $\leq 1.5$ . In such a case, the worse quality of a certain group of data would have to be compensated with a very high quality of the remaining data. The question arises 'how does it impact on the environmental score of the process modelled by the dataset'? Theoretically, it would be important in situations where the poorer quality data would pertain to elements of potentially high environmental relevance. However, the way that the rules for primary data quality rating criteria (Tables 2 and 3) have been defined offer a safeguard [23,26]. These rules do not allow individual data items to exceed certain quality levels (e.g.,  $DQR \leq 3$  for the precision,  $DQR \leq 2$  for the technological representativeness). These levels are restrictive and shall be applied for all data included in the dataset. This means that a differentiation in data quality is possible but only to a limited extent. In this way, the impact on the final environmental score of the dataset is also limited.

Another aspect worth stressing is the question of using emission indicators. From the point of view of data quality, the best-case scenario is when emission data come from direct measurements, although, in practice, this is not always possible (e.g., due to a lack of access to measurement equipment or the specific character of an emission source, such as intestinal fermentation in cattle). In the case of using emission indicators, the emission rate is partially defined on the basis of primary information, coming from the location and time of the process execution (e.g., consumption and specific character of a fuel; a machine's year of manufacture; characteristics of animals; the specific character of feed and environmental conditions of animal breeding). The emission rate is also partially based on the patterns and parameters sourced from secondary sources, which may be subject to modification and subsequent updates of source documents. Thus, the question arises: if or how to make consideration for the geographical, temporal and technological representativeness

of emission indicators sourced from source documents? In our example, we assumed that the year of report publication, as well as the geographical and technological scope of emission indicators contained in the reports, should be considered for the purpose of defining dEFs quality, pertaining to emissions in a farmstead. We have used indicators from the most up-to-date version of the 2019 IPCC reports purposefully to obtain better DQR for the temporal representativeness criterion. If we had used indicators from the 2006 IPCC reports, the temporal offset between the EF report publication date and the age of data on emissions would be too large.

Data and datasets are critical areas, also from the viewpoint of PEF study verification. According to Zampori and Pant [23], the verification and validation of the PEF studies are mandatory whenever the results are used for any type of external communication. The verifier must take into consideration various aspects connected with the data, for example: coverage, precision, completeness, representativeness, consistency, reproducibility, sources and uncertainty, as well as plausibility, quality and accuracy of the LCA-based data [23]. The proposed procedure seems to be easier for reproduction by a verifier and the correctness of the data quality rating may be verified based on a well-documented report. In the case of a simultaneously performed (parallel) verification, a verifier might (on an ongoing basis and without the need to perform LCIA calculations) control the indicators of primary data quality, collected by the study commissioner or their suppliers.

As a supplement to the discussion, we have performed a SWOT analysis for the proposed simplification, the results of which are presented in Table 6.

**Table 6.** Strengths, weaknesses, opportunities and threats of the suggested simplification.

Strengths	Weaknesses
<ul style="list-style-type: none"> <li>■ Criteria for activity data and elementary flows quality rating remain unchanged, pursuant to EF method update [23].</li> <li>■ The company-specific dataset quality rating is very simple and fast.</li> <li>■ Performance of the rating does not require any skills in impact assessment; a little training in rating criteria seems to be sufficient.</li> <li>■ The possibility of independent execution of data quality rating by a data supplier.</li> <li>■ The reproductivity of data quality rating results increases. Meticulous documentation and justification for rating results in a report would greatly facilitate verification.</li> </ul>	<ul style="list-style-type: none"> <li>■ The proposed procedure does not employ the materiality principle, as the dataset quality is not determined with the environmental relevance of individual flows.</li> </ul>
Opportunities	Threats
<ul style="list-style-type: none"> <li>■ Shortening of dataset quality rating performance.</li> <li>■ Data suppliers could control the data quality and manage the process of their collection on their own.</li> <li>■ The possibility of preparing model sheets for collecting data (templates), supporting data quality rating.</li> </ul>	<ul style="list-style-type: none"> <li>■ With a large number of inventory elements, it may be more difficult to obtain the required level of dataset quality.</li> <li>■ The threat of worse quality data pertaining to AD or EFs, with a potentially high environmental relevance; it would be compensated for by high quality elements with less environmental relevance.</li> </ul>

Source: Own elaboration.

## 6. Conclusions

Data quality assessment has been part of life cycle techniques since their very origins. The first works on quality assessment were published in the early 1990s. In many places, the European Commission's methodology for assessing the quality of data and datasets refers to approaches and experiences already developed in the past. Its core lies in a semi-quantitative matrix, conceptually equivalent to the Pedigree Matrix introduced into life cycle techniques by Weidema and Wesnaes [12]. In the framework of the environmental footprints methodology, such matrices were developed to assess the quality of company-specific data (Tables 2 and 3). The Data Needs Matrix, which determines minimum acceptable quality levels for processes, depending on the supervision (control) over the process and access to primary data, constitutes an innovative solution designed by the European Commission, especially for environmental footprints. Each of the PEF studies can be performed at a different time and place, and hence they may differ with respect to initial data needs. Therefore, references for assessing the quality of specific data will differ due to varied temporal, technological and geographical coverage of the product system. The guidelines developed by the European Commission seem to be very effective at managing flexibility (taking into account varying degrees of process control and access to specific data); however, they seem to fail in terms of operability (being cookbook-like). The procedure for identifying significant AD and dEFs in company-specific datasets may be recognised as being complex and time-consuming. The main reason for this is the linking of data and dataset quality rating with impact assessment. Owing to this, however, the materiality approach rule is executed but the procedure of acquiring information, which is the most relevant out of all inventory data, is potentially difficult for people having no experience in LCA/EF analyses. The simplification we propose consists of discarding the use of influence quality ratings in data quality rating, and in adopting the assumption that all AD and dEFs contained in the company-specific dataset are rated. With our example, the procedure has been visibly simplified and it even allows non-LCA experts (e.g., suppliers) to assess and control data quality. The analysis time has been shortened by several times. This is important, especially if many different processes included in the product system are to be modelled with primary data and company specific datasets. The simplification is universal and may be applied for all processes modelled by company-specific datasets. It must be stressed, however, that this is a single-case example, and the proposed approach would require further practical verification. It seems, however, that the proposed solution could contribute to a wider propagation of environmental footprint determination by enterprises, which (and this is a widely known fact) report that these analyses are very complex and require the engagement of specialists with rare competencies.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/en14165004/s1>, Table S1. Assumptions pertaining to the milk-producing farmstead operation that provide company-specific, Table S2: Company-specific data for the production of 168867 FPCM of raw milk in a dairy farm located in Poland.

**Author Contributions:** Conceptualisation, A.L., K.J.-L. and P.K.; methodology, A.L.; software, A.L.; validation, K.J.-L., P.K.; formal analysis, A.L.; resources, A.L., K.J.-L. and P.K.; writing—original draft preparation, A.L.; writing—review and editing, K.J.-L. and P.K.; visualisation, A.L., K.J.-L.; supervision, A.L.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The example presented in this paper refers to the hypothetical case study and is intended to illustrate data quality-related considerations, not to assess the environmental performance of the farm.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

The most important acronyms used in the paper:

AD	Activity data
CSD	Company-specific datasets
DNM	Data Needs Matrix
DQR	Data Quality Rating
dEFs	Direct elementary flows
EFs	Elementary flows
EF	Environmental Footprint
LCA	Environmental Life Cycle Assessment
LCA	Life Cycle Impact Assessment
PEF	Product Environmental Footprint
OEF	Organisation Environmental Footprint
PEFCRs	Product Environmental Footprint Category Rules
OEFSRs	Organisation Environmental Footprint Sector Rules

## References

1. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A New Circular Economy Action Plan for a Cleaner and more Competitive Europe (COM/2020/98). Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN> (accessed on 15 May 2021).
2. Harbi, S.; Margni, M.; Loerincik, Y.; Dettling, J. Life Cycle Management as a Way to Operationalize Sustainability within Organisations. In *Life Cycle Management. Third Volume of LCA-Compendium—The Complete World of Life Cycle Assessment*; Sonnemann, G., Margni, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 23–34.
3. Nygren, J.; Antikainen, R. *Use of Life Cycle Assessment (LCA) in Global Companies*; Reports of The Finnish Environment Institute, 2010. Available online: [https://helda.helsinki.fi/bitstream/handle/10138/39723/SYKEre\\_16\\_2010.pdf](https://helda.helsinki.fi/bitstream/handle/10138/39723/SYKEre_16_2010.pdf) (accessed on 15 May 2021).
4. Feucht, Y.; Zander, K. *Consumers' Attitudes on Carbon Footprint Labelling: Results of the SUSDIEET Project*; Thünen Working Paper (78); Johann Heinrich von Thünen-Institut: Braunschweig, Germany, 2017. Available online: [https://literatur.thuenen.de/digbib\\_extern/dn059137.pdf](https://literatur.thuenen.de/digbib_extern/dn059137.pdf) (accessed on 20 May 2021).
5. Lelek, L.; Kulczycka, J.; Lewandowska, A.; Zarębska, J. Life cycle assessment of energy generation in Poland. *Int. J. Life Cycle Assess.* **2016**, *21*, 1–14. [[CrossRef](#)]
6. Zhang, Y.; Yan, D.; Hu, S.; Gu, S. Modelling of energy consumption and carbon emission from the building construction sector in China, a process-based LCA approach. *Energy Policy* **2019**, *134*, 110949. [[CrossRef](#)]
7. Ciacci, L.; Passarini, F. Life Cycle Assessment (LCA) of Environmental and Energy Systems. *Energies* **2020**, *13*, 5892. [[CrossRef](#)]
8. International Organization for Standardization. *International Standard ISO 14044:2006. Environmental Management—Life Cycle Assessment—Requirements and Guidelines*; International Organization for Standardization: Geneva, Switzerland, 2006.
9. Fava, J.; Jensen, A.A.; Lindfors, L.; Pomper, S.; De Smet, B.; Warren, J.; Vigon, B. (Eds.) *Life-Cycle Assessment Data Quality: A Conceptual Framework*; SETAC Workshop Report; Society of Environmental Toxicology and Chemistry: Wintergreen, VA, USA, 1992.
10. Bakst, J.S.; Lacke, C.J.; Weitz, K.A.; Warren, J.L. *Guidelines for Assessing the Quality of Life Cycle Inventory Analysis*; Report Prepared for Environmental Protection Agency, Research Triangle Institute: Research Triangle, NC, USA, 1995.
11. Vigon, B.W.; Jensen, A. Life cycle Assessment—Data quality and database practitioner survey. *J. Clean. Prod.* **1995**, *3*, 135–141. [[CrossRef](#)]
12. Weidema, B.P.; Wesnaes, M.S. Data quality management for life cycle inventories—an example of using data quality indicators. *J. Clean. Prod.* **1996**, *4*, 167–174. [[CrossRef](#)]
13. Weidema, B.P. Multi-User Test of the Data Quality Matrix for Product Life Cycle Inventory Data. *Int. J. Life Cycle Assess.* **1998**, *3*, 259–265. [[CrossRef](#)]
14. Van den Berg, N.W.; Huppel, G.; Lindeijer, E.W.; Van Der Ven, B.L.; Wrisberg, M.N. *Quality Assessment for LCA*; CML Report 152; Leiden University: Leiden, The Netherlands, 1999.
15. European Commission. *ILCD Handbook. International Reference Life Cycle Data System. General Guide to Life Cycle Assessment—Detailed Guidance*; EUR 24708 EN; Joint Research Centre: Ispra, Italy, 2010. Available online: <https://eplca.jrc.ec.europa.eu/uploads/ILCD-Handbook-General-guide-for-LCA-DETAILED-GUIDANCE-12March2010-ISBN-fin-v1.0-EN.pdf> (accessed on 15 May 2021).
16. Frischknecht, R.; Jungbluth, N.; Althaus, H.-J.; Doka, G.; Dones, R.; Heck, T.; Hellweg, S.; Hirschier, R.; Nemecek, T.; Rebitzer, G.; et al. The Ecoinvent Database: Overview and Methodological Framework. *Int. J. Life Cycle Assess.* **2005**, *10*, 3–9. [[CrossRef](#)]
17. Ciroth, A.; Muller, S.; Weidema, B.P.; Lesage, P. Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *Int. J. Life Cycle Assess.* **2016**, *21*, 1338–1348. [[CrossRef](#)]

18. Lewandowska, A.; Foltynowicz, Z.; Podlesny, A. Comparative LCA of Industrial Objects Part 1: LCA Data Quality Assurance—Sensitivity Analysis and Pedigree Matrix. *Int. J. Life Cycle Assess.* **2004**, *9*, 86–89. [CrossRef]
19. Edelen, A.; Ingwersen, W. *Guidance on Data Quality Assessment for Life Cycle Inventory Data*; U.S. Environmental Protection Agency: Cincinnati, OH, USA, 2016. Available online: [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cmf?Lab=NRMRL&dirEntryId=321834](https://cfpub.epa.gov/si/si_public_record_report.cmf?Lab=NRMRL&dirEntryId=321834) (accessed on 20 May 2021).
20. Edelen, A.; Ingwersen, W.W. The creation, management, and use of data quality information for life cycle Assessment. *Int. J. Life Cycle Assess.* **2018**, *23*, 759–772. [CrossRef] [PubMed]
21. Single Market for Green Products Initiative. Available online: <https://ec.europa.eu/environment/eussd/smgp/index.htm> (accessed on 20 May 2021).
22. Commission Recommendation of 9 April 2013 on the Use of Common Methods to Measure and Communicate the Life Cycle Environmental Performance of Products and Organisations (2013/179/EU). Available online: <https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32013H0179&from=EN> (accessed on 20 May 2021).
23. Zampori, L.; Pant, R. *Suggestions for Updating the Product Environmental Footprint (PEF) Method*; Publications Office of the European Union: Luxembourg, 2019. Available online: <https://ec.europa.eu/jrc/en/publication/suggestions-updating-product-environmental-footprint-pef-method> (accessed on 20 April 2021).
24. Zampori, L.; Pant, R. *Suggestions for Updating the Organisation Environmental Footprint (OEF) Method*; Publications Office of the European Union: Luxembourg, 2019. Available online: <https://ec.europa.eu/jrc/en/publication/suggestions-updating-organisation-environmental-footprint-oef-method> (accessed on 20 April 2020).
25. Product Environmental Footprint Category Rules Guidance. Version 6.3. 2018. Available online: [https://ec.europa.eu/environment/eussd/smgp/pdf/PEFCR\\_guidance\\_v6.3.pdf](https://ec.europa.eu/environment/eussd/smgp/pdf/PEFCR_guidance_v6.3.pdf) (accessed on 20 April 2021).
26. Product Environmental Footprint Category Rules for Dairy Products. Version 1.1. 2020. Available online: [https://ec.europa.eu/environment/eussd/smgp/pdf/PEFCR-DairyProducts\\_Feb%202020.pdf](https://ec.europa.eu/environment/eussd/smgp/pdf/PEFCR-DairyProducts_Feb%202020.pdf) (accessed on 2 April 2021).
27. Gavrilova, O.; Leip, A.; Dong, H.; MacDonald, J.D.; Bravo, C.A.G.; Amon, B.; Rosales, R.B.; Del Pado, A.; De Lima, M.A.; Oyhantcabal, W.; et al. Emissions from Livestock and Manure Management. In *Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*; IPCC: Geneva, Switzerland, 2019; Volume 4, Chapter 10. Available online: <https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4> (accessed on 2 April 2021).
28. De Klein, C.; Novoa, R.S.A.; Ogle, S.; Smith, K.A.; Richette, P.; Wirth, T.C. N<sub>2</sub>O Emissions from Managed Soils, and CO<sub>2</sub> Emissions from Lime and Urea Application. In *Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*; IPCC: Geneva, Switzerland, 2019; Volume 4, Chapter 11. Available online: <https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4> (accessed on 2 April 2021).
29. EMEP/EEA. *Air Pollutant Emission Inventory Guidebook. Technical Guidance to Prepare National Emission Inventories*; EEA Technical Report No 12/2013; EEA: Luxembourg, 2013. Available online: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2013> (accessed on 2 April 2021).
30. EMEP/EEA. *Air Pollutant Emission Inventory Guidebook. Technical Guidance to Prepare National Emission Inventories*; EEA Technical Report No 21/2016; EEA: Luxembourg, 2016. Available online: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2016> (accessed on 2 April 2021).
31. National Centre for Emission Balancing and Management. *Calorific Values and CO<sub>2</sub> Emission Factors in 2016 to Be Reported under the Emission Trading Scheme for 2019*; Institute for Environmental Protection—National Research Institute: Warsaw, Poland, 2018. Available online: [https://www.kobize.pl/uploads/materialy/WO\\_i\\_WE\\_do\\_monitorowania-ETS-2019.pdf](https://www.kobize.pl/uploads/materialy/WO_i_WE_do_monitorowania-ETS-2019.pdf) (accessed on 2 April 2021). (In Polish)
32. National Centre for Emission Balancing and Management. *Calorific Values and CO<sub>2</sub> Emission Factors in 2017 to Be Reported under the Emission Trading Scheme for 2020*; Institute for Environmental Protection—National Research Institute: Warsaw, Poland, 2019. Available online: <https://www.kobize.pl/pl/article/2019/id/1580/wartosci-opalowe-wo-i-wskazniki-emisji-co2-we-w-roku-2017-do-raportowania-w-ramach-systemu-handlu-uprawnieniami-do-emisji-za-rok-2020> (accessed on 2 April 2021). (In Polish)
33. Franze, J.; Baitz, M.; De Schryver, A.; Horlacher, S.; Wolf, M.-A. What Is an EF Compliant Dataset? *Webinar*, 4 June 2019. Available online: [https://ec.europa.eu/environment/eussd/smgp/pdf/webinar\\_what\\_%20is\\_an\\_EF\\_compliant\\_dataset.pdf](https://ec.europa.eu/environment/eussd/smgp/pdf/webinar_what_%20is_an_EF_compliant_dataset.pdf) (accessed on 11 July 2021).
34. Annex C List of Default CFF Parameters. Available online: <http://eplca.jrc.ec.europa.eu/lcdn/developeref.xhtml> (accessed on 2 April 2021).