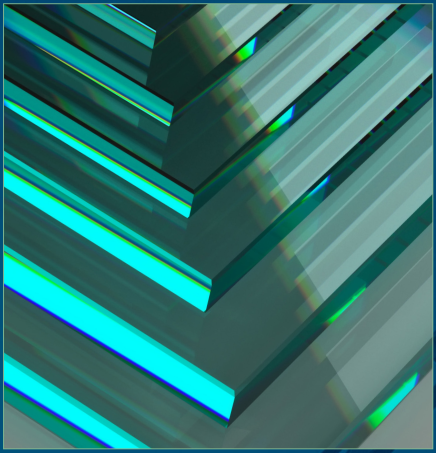

Data Quality Assessment Method to Support the Label Program for Low Embodied Carbon Construction Materials (Version 1)

Office of Chemical Safety and Pollution Prevention
August 2024

Table of Contents

Tables

Figures

Background

This document supports Phase I of the label program: data quality improvement. The overall purpose of this phase is to ensure underlying data used for the label program are suitable for use for procurement.

Specific use cases of the data quality assessment method include the following:

- Life cycle inventory data developers can apply the method to score processes and flows when generating new datasets.
- Product category rule committees can use the DQA results to support prescribing secondary datasets.
- Life cycle assessment modelers can use the DQA results to support selection of secondary LCI datasets.
- LCA and EPD developers can use the method to provide data quality information for their foreground LCI data.

The objective of this DQA method is to provide a systematic approach primarily for evaluating the quality of secondary LCI datasets used in developing LCAs produced for PCRs, and secondarily in developing LCAs for environmental product declarations. This DQA method can be applied to both primary and secondary data during the development of LCAs produced for PCRs and for LCAs produced for EPDs. This DQA method can also be used to help improve the Federal LCA Commons[1] by evaluating current data on the FLCAC to identify deficiencies and gaps.

EPA's *Vision and Plan to Improve Secondary Life Cycle Assessment Data Used in Environmental Product Declarations* provides more information on secondary data development efforts for improving the quality and quantity of datasets on the FLCAC. Meanwhile, *U.S. EPA Criteria for Product Category Rules (PCRs) to Support the Label Program for Low Embodied Carbon Construction Materials* provides further guidance on the requirements for LCI data quality scoring for LCAs used for PCRs. In EPA's PCR Criteria, application of EPA's DQA method is a leadership criterion, and PCRs are encouraged but not required to apply this DQA method for the label program.

A companion Excel template for completing DQAs according to this method is also available. The companion template includes an example case study showing how to implement this DQA method. If a PCR Committee would like EPA support, EPA can conduct this DQA method for construction materials that are a part of EPA's label program, provided that the LCA used for the PCR is made available to EPA, and provided that the PCR covers the United States and uses ISO 21930:2017 as its core PCR document. Such support is subject to available resources.

This is version 1 of the DQA method; EPA may periodically update the document as more data becomes available. EPA may also periodically check and meet with sector-specific stakeholders about updating and implementing the method as necessary based on updates (or planned updates) to key documents, such as PCRs. Requests for EPA support in conducting this DQA method and feedback or questions on implementation of this DQA method can be submitted through the embodiedcarbon@epa.gov email.

---

[1] The Federal LCA Commons is a federal inter-agency initiative to provide free-to-use and publicly accessible data for use in LCAs.

This DQA method should be cited as:

U.S. EPA. (2024). Data Quality Assessment Method to Support the Label Program for Low Embodied Carbon Construction Materials (Version 1).

Approach

EPA drew on multiple existing sources to develop a suitable DQA method for the label program. The method broadly encompasses multiple characteristics of data quality identified across the reviewed sources, including LCI data quality indicators, suitability for life cycle impact assessment methods relevant for the label program, and more generalized important data attributes. The following sources were referenced in development of the DQA method:

- EPA's LCI DQA method (Edelen and Ingwersen, 2016).
- Methods used by the Federal Highway Administration in published LCI tool development including the Pave tool (FHWA, 2021).
- Methods used in the European Union Product Environmental Footprint program to evaluate secondary data quality (Fazio et al., 2020).
- The American Center for Life Cycle Assessment's *Guidance for Assessing Data Quality of Background Life Cycle Inventory (LCI) Datasets* (ACLCA, 2022).
- Characteristics of fit for purpose secondary data defined by the Interagency Team on Secondary Data for EPDs. (More information on the Interagency Team is provided in the companion document *Vision and Plan to Improve Secondary Life Cycle Assessment Data Used in Environmental Product Declarations*.)
- Other draft approaches for DQA provided by the Interagency Team on Secondary Data for EPDs.
- DQA methods used or in development by PCR committees.
- ISO standards 21930:2017 and 14044:2006 (ISO, 2017, 2006).

Additionally, this DQA method takes into account *U.S. EPA Criteria for Product Category Rules (PCRs) to Support the Label Program for Low Embodied Carbon Construction Materials* —the PCR Criteria document—which defines requirements of LCI datasets. Key requirements of EPA's PCR Criteria relevant to this DQA method include:

- Criterion 3.1.B: The PCR shall clearly specify the scope and data quality for secondary data and include recommendations for free-to-use and publicly accessible datasets or databases facilitating this process.
- Criterion 3.2.A: Effective January 1, 2026, the LCA(s) used for the PCR shall include a complete DQA for both primary and secondary data, including the specification of which DQA method was used.
- Criterion 3.2.B: Specific data (i.e., from upstream EPDs) that are representative of the raw material supply chain shall be used where possible. Where specific data are not possible, PCRs shall prescribe free-to-use and publicly accessible secondary datasets. PCRs shall prescribe a unique free-to-use and publicly accessible secondary dataset for each of the following flows:

  - Electricity
  - Fuels
  - Transportation
  - Other unit processes in which secondary data are required by the PCR

  Note: Effective January 1, 2026, PCRs shall prescribe the use of the EPA-designated free-to-use and publicly accessible datasets for the flows identified within this criterion. Prior to this date, PCRs that are being updated shall provide a commitment to use public datasets in the future if they are not already using public data. If a PCR uses private datasets, the PCR shall outline why public datasets are not adequate for the flows the PCR is seeking to model.

- Criterion 2.1.D: Requires inclusion of LCIA categories outlined in Table 5 of ISO 21930:2017, which is relevant to assessing completeness of an LCI dataset for the label program. These include the following LCIA categories and all associated elementary flows, as defined in their source methods: greenhouse gas, ozone depletion, eutrophication, acidification and photochemical oxidant creation.[2]

Additionally, in January 2023, EPA released a formal public Request for Information associated with the label program and received over 100 responses that addressed the importance of the quality and quantity of secondary data. RFI feedback included a recommendation for EPA provision of a standardized DQA method for its label program such as an enhanced version of the existing (2016) EPA's LCI DQA method. In response to these public comments, the Interagency Team on Secondary Data for EPDs identified attributes of fit for purpose secondary data. Many of the identified attributes align with data quality indicators identified in the other reviewed data sources. However, some important data attributes identified by the Interagency Team on Secondary Data for EPDs—for example, "transparent," "reproducible," "interoperable," "maintained," "publicly available"—are important for the EPA label program but are not traditional LCA DQIs.[3] This DQA method also incorporates these important attributes.

Data Quality Indicators

To develop the label program DQA method, EPA extracted DQIs from the reviewed sources, which generally assessed the same types of DQIs. These DQIs were compiled and further refined based on the needs of the label program. Development of the approach for each DQI is described in the subsequent sections.

Indicators are defined for important data attributes and at the flow level or process level. Flow-level DQIs should be defined and assessed for elementary flows and technosphere flows within each unit process, including both technosphere input and output flows. Process-level DQI and data attributes should be defined at the overall process level. For each indicator, the best score is 1, while a score of 5 represents the lowest data quality (default to 5 if unknown). A unique indicator for assessing the process's compatibility with LCIA categories, as defined in Table 5 of ISO 21930:2017, is also included.

---

[2] EPA has outlined LCIA indicators for EPDs and PCRs (available on EPA's website). These are the preferred factors to use until an update to the Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (TRACI 3.0) is released.

[3] Traditional LCI DQIs include temporal, geographical and technological representativeness.

Flow-Level DQIs

Table 1 describes the flow-level DQIs to be assessed under this LEC DQA method. This aligns with the EPA LCI DQA (previously developed in 2016). The existing EPA LCI DQA, FHWA Pave DQA and EU PEF data quality rating methods all consider four different aspects of representativeness (temporal, geographical, technological, data collection methods). Seasonality may be important for the construction material supply chain of certain materials and shall be assessed by the "data collection methods" indicator. Reliability is also included as a DQI at the flow level. Reliability indicates quality of data generation method and verification of data collection methods.

Table 1. Flow-Level DQIs

Indicator	Definition	← Highest Data Quality (Lowest Score)			Lowest Data Quality (Highest Score) →	
		1	2	3	4	5 (Default)
Temporal representa-tiveness	Indicates the temporal difference between the date of data generation and the date the data are supposed to represent based on the PCR.	Less than 3 years of difference	Less than 6 years of difference	Less than 10 years of difference	Less than 15 years of difference	Age of data unknown or more than 15 years
Geographic-al representa-tiveness	Indicates how well the geographical area from which data for a unit process are collected satisfies the goal of the study (ISO 14044).	Data from same resolution[a] and same area of study	Within one level of resolution and a related area of study[b]	Within two levels of resolution and a related area of study	Outside two levels of resolution but a related area of study	From a different or unknown area of study
Technolog-ical representa-tiveness	Indicates technical representativeness based on four categories: process design, operating conditions, material quality/type and process scale.	All technology categories[c] are equivalent	Three of the technology categories are equivalent	Two of the technology categories are equivalent	One of the technology categories is equivalent	None of the technology categories are equivalent
Data collection methods	Assessment of the robustness of the sampling methods and data collection period.	Representa-tive data from >80% of the relevant market,[d] over an adequate period[e]	Representa-tive data from 60–79% of the relevant market, over an adequate period, or representa-tive data from >80% of the relevant market, over a shorter period	Representa-tive data from 40– 59% of the relevant market, over an adequate period, or representa-tive data from 60–79% of the relevant market, over a shorter period	Representa-tive data from <40% of the relevant market, over an adequate period, or representa-tive data from 40–59% of the relevant market, over a shorter period	Unknown or data from a small number of sites and from shorter periods

Indicator	Definition	← Highest Data Quality (Lowest Score)			Lowest Data Quality (Highest Score) →	
		1	2	3	4	5 (Default)
Reliability	Indicates quality of data generation method and verification of data collection methods.	Verified[f] data based on measurements	Verified data based on a calculation or non-verified data based on measurements	Non-verified data based on a calculation	Documented estimate	Undocumented estimate

[a] Levels of resolution are defined as follows: global, continental, sub-region, national, state/province/region, county/city, site-specific. (The first four of these are from the UN geoscheme [United Nations, 2013].) The same approach applies for imports to the United States.

[b] A related area of study is defined by the user and should be documented in the geographical metadata. By default, a related area of study is one within the same hierarchy of political boundaries (e.g., Denver is within Colorado, which is within the United States, which is within North America).

[c] Technology categories are process design, operating conditions, material quality and process scale.

[d] The relevant market should be documented in the DQA. The default relevant market is measured in production units; if the relevant market is determined using other units, this should be documented in the DQA. The relevant market established in the metadata should be consistently applied to all flows within the unit process.

[e] An adequate time period can be evaluated as one long enough to even out normal fluctuations. The default period is one year, except for emerging technologies (two to six months) or agricultural projects over three years. Seasonality considerations shall be incorporated for construction materials where relevant.

[f] Verification may take place in several ways, e.g. by on-site checking, by recalculation, through mass balances or crosschecks with other sources. For values calculated from a mass balance or another verification method, an independent verification method must be used in order to qualify the value as verified.

Use Case and Relative Scoring

The three representativeness DQI (temporal, geographical and technological) are dynamic indicators, requiring a comparison to be completed before a data quality score is assigned; data collection methods and reliability are static indicators, meaning they are defined by the methods used to complete and communicate the study. The dynamic indicators require definition of the physical system the data should represent, or the use case. This use case must be defined to determine the dynamic DQI scoring for the data being evaluated.

For example, consider a set of hypothetical datasets named Model A and Model B that are intended to simulate the same product. The difference between these two models is that Model A is composed of production data based on averages from one U.S. state, while Model B is composed of production data based on averages from the entire United States. The resulting geographical representativeness scores will be different depending on whether the data are intended to be used to represent a process within the specific state Model A is based on or intended to be used to represent the entire U.S. production. Herein lies the critical value of defining a use case. The representativeness DQIs can only be evaluated relative to a specific application.

Table 2 lists information required to be collected for a use case before completing the flow-level representativeness data quality scoring.

Table 2. Use Case DQA Inputs

Representativeness Category	DQI Component	Use Case Value
Temporal	Year of data use	Data entry
Geographical	Geographical resolution	Data entry
Technological	Process design	Data entry
	Operating conditions	Data entry
	Material quality/type	Data entry
	Process scale	Data entry

Identifying the Flow Value to Score

This DQA method applies flow-level DQIs to both elementary and technosphere flows. The value that should be assessed for scoring is the direct quantity listed as an input or output. For example, if there is an input of 10 tonne-kilometers of truck transport to a petroleum refining process, how the value of "10" was derived should be scored, not that entire "truck transport" input process. The data quality implications of the upstream truck transport process will be assessed in the data quality scoring for the truck transport process itself, not in the data quality scoring for the petroleum refining process. More details on aggregating flow-level DQIs are provided in subsequent sections of this document "Flow-Level DQI Aggregation". If a flow value is a combination of multiple parameters (e.g., emission factor X activity factor), the flow data quality score should use the lowest data quality (highest score) of the input parameters.

Process-Level DQI

Table 3 provides the DQIs for processes; in this context, a process is composed of other flows such as elementary and technosphere flows—inputs and outputs. These DQIs draw on process-level DQIs from the EPA LCI DQA and the FHWA Pave DQA.

Table 3. Process-Level DQIs

Indicator	Definition	← Highest Data Quality (Lowest Score)		Lowest Data Quality (Highest Score) →		
		1	2	3	4	5 (Default)
Reviewed	Assesses whether data have been independently quality assured and reviewed by subject matter and LCA experts (types of reviewers).	Documented reviews by at least two types of third-party reviewers	Documented reviews by at least two types of reviewers, one a third party	Documented review by a third-party reviewer	Documented review by an internal reviewer	No documented review
Flow completeness	Indicates how well all of the flows intended by the study system boundary are captured (e.g., as intended by PCR system diagram). Flows include resource, material, energy and water inputs and product, emission, discharge, waste outputs.	>80% of determined flows have been evaluated and given values	60–79% of determined flows have been evaluated and given values	40–59% of determined flows have been evaluated and given values	<40% of determined flows have been evaluated and given values	Flow completeness unknown based on available metadata

Indicator	Definition	← Highest Data Quality (Lowest Score)			Lowest Data Quality (Highest Score) →	
		1	2	3	4	5 (Default)
Range data complete-ness	Indicates inclusion of range data for flows. This should be based on quantitative metadata and not a distribution based on data quality. This should preferably be a calculated metric of the spread of the data (mean and standard deviation), or an estimate of a triangular distribution in the case of three data points, or a uniform distribution in the case of just minimum and maximum data.	>80% of determined flows include range data	60–79% of determined flows include range data	40–59% of determined flows include range data	<40% of determined flows include range data	Range data not available for any flows or unknown

Data Attributes

The Table 4 data attributes were identified as important by the Interagency Team. For this DQA method, considerations for fit for purpose secondary data attributes take a broader view aimed at covering important secondary data attributes for the label program, rather than more narrow definitions of data quality used traditionally in LCA (ISO 14044:2006).

Table 4. Data Attributes to Be Assessed at the Process Level

Indicator	Definition	← Highest Data Quality (Lowest Score)			Lowest Data Quality (Highest Score) →	
		1	2	3	4	5 (Default)
Reproduci-bility	Indicates transparency of the underlying model. Assesses how well sources and calculations of the underlying model are documented to enable a third party to independently recreate the background data result based on the data documentation (metadata).	Underlying model and associated calculations are fully transparent and based on public sources (e.g., source code underlying modeling is linked and model is free-to-use and publicly accessible)	Underlying model and associated calculations are fully transparent and based on non-public sources	Underlying model and associated calculations are partially transparent and based on public sources	Underlying model and associated calculations are partially transparent and based on non-public sources	Underlying model and associated calculations are not transparent and based on non-public sources or unknown

Indicator	Definition	← Highest Data Quality (Lowest Score)			Lowest Data Quality (Highest Score) →	
		1	2	3	4	5 (Default)
Free-to-use and publicly accessible	Indicates whether the LCI data are free-to-use and publicly accessible, along with the level of LCI data aggregation.	Full linked LCI unit process data are free-to-use and publicly accessible (including background data inputs)	Full linked LCI unit process data are free to-use and publicly accessible (some background data inputs based on proprietary inputs)	Full LCI system process data are free-to-use and publicly accessible	Non-public LCI unit process data	Non-public LCI system process data or unknown
Interoper-able	Indicates existence of data structure and nomenclature that enable utilization with external LCA datasets/software.	Foreground and input background data comply with FLCAC nomenclature and data structure requirements[a]	Foreground and input background data comply with other recognized data and nomenclature structure such as ILCD	At least foreground data comply with FLCAC nomenclature and data structure requirements[a]	At least foreground data comply with other recognized data and nomenclature structure such as ILCD	Data do not adhere to established data structure and nomencla-ture system
Maintained	Indicates existence of long-term resources (funding) and update plans to support the dataset.	Plans and resources for future updates are present and communicated; updates occur at least annually	Plans and resources for future updates are present and communicated; updates occur at least every three years	Plans and resources for future updates are present and communicated; updates occur at least every five years	Plans and resources for future updates are present but resource availability is unknown	Plans and resources for future updates are not present or unknown

[a] Dataset aligned with the FLCAC elementary flow list and technosphere flows have been attached to default providers. JSON-LD format available.

LCIA Method Compatibility

The LCIA method compatibility indicator is determined based on a process's quantified flows for five impact categories outlined in Table 5 of ISO 21930:2017. Table 5 provides a scoring table for calculating the final indicator score. The final score is the average of five assigned scores for the five impact categories. Because each impact category can be scored 1, 3, or 5, the final score of the LCIA method compatibility indicator can range from 1 to 5. To determine how to score LCIA method compatibility, the LCA documentation associated with the dataset should be assessed to determine if the data were generated to support specific LCIA methods/categories. Elementary flows present in the dataset should be compared against EPA's LCIA method to determine compatibility with each LCIA category.[4] The results from the process-level flow completeness data quality scoring can also be used for assessing

---

[4] EPA has outlined LCIA indicators for EPDs and PCRs (available on EPA's website). These are the preferred factors to use until an update to TRACI 3.0 is released.

LCIA method compatibility (e.g., to determine if a specific flow relevant to LCIA categories is expected to be present for the process being evaluated).

Table 5. LCIA Method Compatibility Indicator Scoring Table

Impact Category	Identified Compatibility with Impact Assessment Category[a]	Known Flow Missing[b]	Not Compatible or unknown
GHGs	1	3	5
Ozone depletion potential	1	3	5
Eutrophication potential	1	3	5
Acidification potential	1	3	5
Photochemical oxidant creation potential	1	3	5

[a] Metadata associated with the dataset indicate it was developed for compatibility with specified LCIA categories.
[b] An example of a case in which flows are missing is when only criteria air pollutants are assessed for a process. Some impact assessment categories, such as photochemical oxidant creation, include some of the six criteria air pollutants (e.g., carbon monoxide, nitrogen oxides), but the photochemical oxidant creation category also characterizes elementary flows beyond these criteria air pollutants. The process should be assessed to determine if elementary flows relevant to photochemical oxidant creation beyond criteria air pollutants are expected to be present.

Flow-Level DQI Aggregation

A weighted average approach to determine flow-level DQI scores should be applied to calculate the aggregated flow-level DQIs for a process. This is a similar approach to the one available in the openLCA software, where the existing (2016) EPA LCI DQA method is available for use.[5] This EPA DQA method uses impact results (in this case, for GHGs) as the weighting factor for both elementary flows and technosphere flows in a process. The impact results are the results of the flow amount and its corresponding characterization factor value. The calculation is presented in Equation 1:

$$\frac{\sum_i Q_i E_i}{\sum_i E_i} \qquad \text{Equation 1}$$

where *Q* represents the data quality of the direct technosphere or elementary flow (*i*) in the inventory and *E* represents result of the upstream or direct impacts (in this case, GHG emissions) of the flow *i*.

When data quality is unknown, a default of 5 should be applied to ensure a complete DQA. Given the time intensity of applying flow-level DQI, this EPA DQA method is only intended to be applied for flows that influence total GHG emissions present in EPA's LCIA method.[6,7]

Figure 1 shows a hypothetical flow diagram of the production of 1 kg of petroleum product; the mock inventory data are presented in Table 6. The process has two technosphere flow inputs and one elementary flow output. Each of the flows has its own data quality score.

---

[5] Source code for this aggregation method is available at https://github.com/GreenDelta/olca-modules/blob/master/olca-core/src/main/java/org/openlca/core/math/data_quality/DQResult.java#L89.
[6] EPA has outlined LCIA indicators for EPDs and PCRs (available on EPA's website). These are the preferred factors to use until an update to TRACI 3.0 is released.
[7] At minimum, the flow-level DQI should be applied for carbon dioxide, methane and nitrous oxide elementary flows and technosphere exchange quantities that may influence the quantity of these flows.

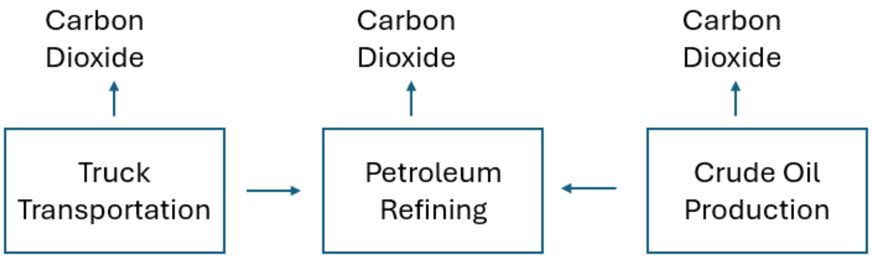

Figure 1: Mock Flow Diagram.

Table 6. Mock Unit Process Flow-Level DQI Aggregation

Process 1: Petroleum Refining (1 kg)				
Input/Output	**Type**	**Amount**	**Unit**	**Mock Data Quality Score (Use Temporal Representativeness as an Example)**
Truck transportation	Technosphere flow	10	tkm	3[a]
Crude oil production	Technosphere flow	1.5	kg	2[a]
Carbon dioxide	Elementary flow (emission)	4	g	4[a]
Process 2: Truck Transportation (1 tkm)				
Input/Output	**Type**	**Amount**	**Unit**	**Mock Data Quality Score (Use Temporal Representativeness as an Example)**
Carbon dioxide	Elementary flow (emission)	8	g	1[b]
Process 3: Crude Oil Production (1 kg)				
Input/Output	**Type**	**Amount**	**Unit**	**Mock Data Quality Score (Use Temporal Representativeness as an Example)**
Carbon dioxide	Elementary flow (emission)	12	g	4[b]

[a] Data quality score should be applied for the exchange value for technosphere flows (e.g., 10 tkm), not for the data quality of the input dataset (e.g., truck transport).

[b] Upstream elementary flow data quality scores are not included in the data quality calculation for the process under evaluation.

The three inputs/outputs for petroleum refining are truck transportation, crude oil production and direct carbon dioxide emissions.

The GHG results for the three inputs/outputs are:[8]

$10\ \text{tkm} \times 8\ \text{g/tkm} \times 1\ \text{kg CO}_2\text{eq/kg CO}_2 = 80\ \text{g CO}_2\text{eq}$ (for truck transportation)
$1.5\ \text{kg} \times 12\ \text{g/kg} \times 1\ \text{kg CO}_2\text{eq/kg CO}_2 = 18\ \text{g CO}_2\text{eq}$ (for crude oil production)
$1\ \text{kg} \times 4\ \text{g/kg} \times 1\ \text{kg CO}_2\text{eq/kg CO}_2 = 4\ \text{g CO}_2\text{eq}$ (for the direct CO_2 emission for petroleum refining)

---

[8] In this case, the carbon dioxide emissions are converted to GHG impact using the characterization factor value: 1 kg CO_2eq/1 kg CO_2.

Using Equation 1, the final data quality score (temporal representativeness as an example) for the petroleum refinery process is equal to:

$$\frac{(3 \times 80\ \mathrm{g\ CO_2eq} + 2 \times 18\ \mathrm{g\ CO_2eq} + 4 \times 4\ \mathrm{g\ CO_2eq})}{(80\ \mathrm{g\ CO_2eq} + 18\ \mathrm{g\ CO_2eq} + 4\ \mathrm{g\ CO_2eq})} = 3$$

From each input/output in the main process, the DQA method calculates the CO_2eq of that input/output and multiplies it by the data quality score of that exchange (for technosphere flows). Elementary flows simply use the direct emissions from the main process.

Resulting DQA Matrix

DQI scores should then be aggregated in the following template (Table 7) to assess the overall data quality for each dataset under investigation. The lowest sum product score indicates the highest data quality. The last two columns in Table 7 should be repeated to compare the data quality of multiple datasets covering similar processes or materials. In version 1 of this DQA, the same priority weighting factor is applied to each indicator. This may be periodically updated by EPA.

Table 7. DQA Matrix

Indicator		Priority Weighting Factor (Lower Number = Higher Priority Weighting)	*[Insert full dataset name, including information on database and version]* *Repeat columns to assess additional datasets*	
			Aggregated DQI Score (Between 1 and 5)	Aggregated DQI Score × Priority Weighting Factor
Aggregated Flow-Level DQI *(to be assessed for flows influencing GHG)*	Temporal representativeness	1	Data entry	Calculation
	Geographical representativeness	1	Data entry	Calculation
	Technological representativeness	1	Data entry	Calculation
	Data collection methods	1	Data entry	Calculation
	Reliability	1	Data entry	Calculation
Process-Level DQI	Reviewed	1	Data entry	Calculation
	Flow completeness	1	Data entry	Calculation
	Range data completeness	1	Data entry	Calculation
Data Attributes	Reproducibility	1	Data entry	Calculation
	Free-to-use and publicly accessible	1	Data entry	Calculation

Indicator		Priority Weighting Factor (Lower Number = Higher Priority Weighting)	*[Insert full dataset name, including information on database and version]* *Repeat columns to assess additional datasets*	
			Aggregated DQI Score (Between 1 and 5)	Aggregated DQI Score × Priority Weighting Factor
	Interoperable	1	Data entry	Calculation
	Maintained	1	Data entry	Calculation
LCIA Method Compatibility		1	Data entry	Calculation
			Total	Calculation (sum)

Dataset DQA scores shall only be compared if the boundaries of each dataset are similar. For example, a total DQA score for gate-to-gate petroleum refining shall not be compared directly to a cradle-to-gate petroleum refining dataset (which also includes upstream life cycle stages such as crude oil production). To align boundaries for compared datasets, upstream processes that contribute to 10 percent or more of the GHG emissions based on the available EPA LCIA method[9] shall also be assessed for data quality. For a dataset where multiple unit processes need to be assessed, Table 7 shall be completed (at minimum) for each process that meets this 10 percent cut-off criterion. It is preferable to complete the data quality scoring for processes below this cut-off criterion, but may be too time intensive initially (until implementation of this DQA method is more widespread). Table 8 displays an example of deriving weights to use when combining DQA results for multiple processes. The weights from the second column in Table 8 shall be used to combine multiple instances of Table 7 for connected processes. Note that currently the weights in Table 8 are based on only the direct child process of the main parent process under evaluation.

---

[9] EPA has outlined LCIA indicators for EPDs and PCRs (available on EPA's website). These are the preferred factors to use until an update to TRACI 3.0 is released.

Table 8. Example of Determining Weights for Dataset with Multiple Processes

Contribution (%)	For Weighted Average DQA (%)	Process	Required Amount	Unit	Total Result (kg CO_2eq)	Note
100	100 - 43 = 57	Petroleum refining, at refinery, U.S.	0.000252	m^3	0.19	This represents the sum of the cradle-to-gate life cycle impacts
43	43	Crude oil, production mixture, at extraction, U.S.	0.21	kg	0.08	This represents the crude oil and upstream impacts

References

ACLCA. (2022). Guidance for Assessing Data Quality of Background Life Cycle Inventory (LCI) Datasets. https://aclca.org/wp-content/uploads/2022-ACLCA-PCR-Guidance_Addendum_Assessing_LCI_Quality_05252022-1.pdf.

Edelen, A., & Ingwersen, W. (2016). Guidance on Data Quality Assessment for Life Cycle Inventory Data. U.S. Environmental Protection Agency. EPA/600/R-16/096. https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryId=321834

Fazio, S., Zampori, L., Schryver, A.D., Kusche, O., Thellier, L., & Diaconu, E. (2020). Guide for EF Compliant Data Sets. Publications Office of the European Union. https://doi.org/10.2760/537292.

FHWA. (2021). LCA Pave: A Tool to Assess Environmental Impacts of Pavement Material and Design Decisions: Underlying Methodology and Assumptions. U.S. Department of Transportation. https://www.fhwa.dot.gov/pavement/lcatool/LCA_Pave_Tool_Methodology.pdf.

GreenDelta. (2024). Data Quality. In: openLCA 2 Manual. https://greendelta.github.io/openLCA2-manual/advanced_top/data_quality.html.

ISO. (2006). ISO 14044:2006—Environmental Management—Life Cycle Assessment—Requirements and Guidelines. https://www.iso.org/standard/38498.html.

ISO. (2017). ISO 21930:2017—Sustainability in Buildings and Civil Engineering Works—Core Rules for Environmental Product Declarations of Construction Products and Services. https://www.iso.org/standard/61694.html.

Kahn, E., Antognoli, E., & Arbuckle, P. (2022). The LCA Commons—How an Open-Source Repository for US Federal Life Cycle Assessment (LCA) Data Products Advances Inter-agency Coordination. Applied Sciences, 12(2), 865. https://doi.org/10.3390/app12020865.

United Nations. (2013). Methodology: Standard Country or Area Codes for Statistical Use (M49). http://unstats.un.org/unsd/methods/m49/m49regin.htm.

USDA. (n.d.). Federal LCA Commons. https://www.lcacommons.gov.

U.S. EPA. (2024a). Characterization Factors for Construction Material EPD Indicators (ISO21930-LCIA-US) v0.1. https://catalog.data.gov/dataset/characterization-factors-for-construction-material-epd-indicators-iso21930-lcia-us-v0-1.

U.S. EPA. (2024b). Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts (TRACI). https://www.epa.gov/chemical-research/tool-reduction-and-assessment-chemicals-and-other-environmental-impacts-traci.

Terminology

Background data: Data contained within the process(es) supporting the foreground system. Includes energy and materials that are delivered to the foreground system as aggregated datasets, in which individual plants and operations are not identified.

Environmental product declaration (EPD): An environmental claim providing quantified environmental data using predetermined parameters and, where relevant, additional environmental information. An EPD also includes additional product and company information. This definition is consistent with the one in ISO 14025:2006.

Foreground data: Data contained within the process(es) a manufacturer is modeling for its product system.

Life cycle assessment (LCA): Refers to the compilation and evaluation of the inputs, outputs and potential environmental impacts of a product system throughout its life cycle. This definition is consistent with the one in ISO 14044:2006.

Life cycle impact assessment (LCIA): Refers to the phase of LCA aimed at understanding and evaluating the magnitude and significance of the potential environmental impacts for a product system throughout the life cycle. This definition is consistent with the one in ISO 21930:2017.

Life cycle inventory (LCI): Phase of life cycle assessment involving the compilation and qualification of inputs and outputs for a product throughout its life cycle. This definition is consistent with the one in ISO 14044:2006.

Primary data: Data determined by direct measurement, estimation or calculation based on specific original source measurements for the specific system under investigation. This definition is based upon the one in ISO 21930:2017.

Product category rules (PCRs): A set of specific rules, requirements and guidelines for developing EPDs for one or more product categories. This definition is consistent with the one in ISO 14025:2006.

Secondary data: Data indirectly determined through measurement, estimation or calculation and not based on specific original source measurements. This can include data that are originally developed using primary data sources, but are further aggregated to represent average processes or products. This definition is based on the one in ISO 21930:2017.

List of Abbreviations

ACLCA	American Center for Life Cycle Assessment
CO_2eq	carbon dioxide equivalent
DQA	data quality assessment
DQI	data quality indicator
EPA	U.S. Environmental Protection Agency
EPD	environmental product declaration
EU	European Union
FHWA	Federal Highway Administration
FLCAC	Federal Life Cycle Assessment Commons
GHG	greenhouse gas
ILCD	International Life Cycle Data
ISO	International Organization for Standardization
LCA	life cycle assessment
LCI	life cycle inventory
LCIA	life cycle impact assessment
PCR	product category rule
PEF	Product Environmental Footprint
RFI	request for information
tkm	tonne-kilometer
TRACI	Tool for Reduction and Assessment of Chemicals and Other Environmental Impacts
USDA	U.S. Department of Agriculture