Guidance for

Assessing Data Quality of Background Life Cycle Inventory (LCI) Datasets

Authors

- **Amlan Mukherjee**, Professor, Michigan Tech Civil, Environmental, and Geospatial Engineering
- **Chaitanya Ganesh Bhat**, Ph.D., LCACP | Sustainability Engineer, Asphalt Institute *(Previously Research Civil Engineer, Engineering & Software Consultants, Turner Fairbanks Highway Research Center)*

Problem

Data quality assurance guidance is outlined in ISO Standards and can be used to check for the quality of the datasets being used. However, there is limited guidance regarding the choice of background Life Cycle Inventory (LCI) data. This can lead to the development of Life Cycle Assessment (LCA) studies supporting Environmental Product Declarations (EPDs) that may have incomparable outcomes (Ingwersen and Stevenson, 2012). Sometimes, even with the same foreground data, a choice of different background inventories can create significant differences in the final LCA results (Willis, 2015).

Solution

The specific recommendations for the choice of background LCI datasets are as follows:

1. The program operators (Pos) **should** adopt the Enhanced Pedigree Matrix (EPM), provided in the Resource section of this document, for consistent data quality assessment.
2. The Pos, in consultation with the PCR Committee, **should** identify and prescribe the most suitable, publicly available background LCI datasets for their products, as determined by use of the EPM, limiting EPD developers to only select foreground data.
3. For programs functioning in the United States, the Pos, in consultation with the PCR Committee, **should** prescribe the LCIs for fossil fuels, transportation, equipment, and electricity provided in the Supporting Resources section of this document.
4. When suitable background LCI datasets are unavailable, Pos, in consultation with the PCR Committee, **should** avoid the use of proxy LCI and instead candidly declare data gaps. If such a data gap is significant, Pos should collaborate with partner industries to develop an LCI that can fill the gap. In the event they exist, the significance of data gaps **shall** be defined by the PCR Committee (e.g., greater than 1% by mass).
5. Pos, in consultation with the PCR Committee, **shall** provide justification for any deviations in the PCR to the above recommendations.

Basis for recommendation

It is suggested that POs adopt a common approach to assessing data quality, and when possible, specifically prescribe background datasets in the PCR, including requirements for regionalization. The benefits of this prescription include:

- EPDs conformant with the PCR will be comparable, as the reported mid-point indicators will only reflect the differences in the facility- and product-specific foreground data.
- Authorities requiring EPDs can have a higher level of confidence in the EPD outcomes, as the data quality assessment for background data has already been completed by the PO.
- POs can afford more transparency into the EPD development process while also removing barriers to third parties accessing the processes and calculations supporting the underlying LCA.

Of course, this does not imply that proprietary data cannot be used for background processes. Where proprietary data is involved, system level inventories can be prescribed, as they obscure unit process information while providing detailed input-output inventories useful for LCA calculations. An example of this is the system level inventories developed by Sphera for the cradle-to-terminal impacts of asphalt binder, which were made available for use by downstream products such as asphalt mixtures. *See the Resource section.*

POs of PCRs that support products in related supply chains could collaboratively choose to prescribe common background inventories. In doing so, they can ensure the ability for downstream products to use system level inventories based on EPDs of upstream products as background data.

Resource: Enhanced Pedigree Matrix

This section outlines the recommended data quality assessment methodology using the Enhanced Pedigree Matrix (Bhat 2020). A pragmatic desired data quality is defined for each category **(indicated in blue)**, and limitations to reach this desired data quality are assessed for different background data categories. This methodology and assessment were completed for the background data recommended for the FHWA Pavement Framework and used in the *LCA Pave* tool, ensuring consistency with the Pavement LCA Framework.

Flow level

Flow level assessment enables evaluation of metadata associated with both product flows and elementary flows, such as name, unit, CAS number, and molecular formula.

Reliability of the data

Reliability is assessed at the flow level and indicates the methods used to generate the data and verification/validation of these methods. In order to point at the specifics of the data collection methods and their validation, the pavement-specific pedigree matrix details four questions within the reliability criterion, and the data quality assessment is carried out as follows:

a) Is the inventory data checked for mass/energy balance, recalculation, etc.?

i) Verified data based on measurements – score 1
ii) Verified data based on a calculation, or on non-verified data based on measurements – score 2
iii) Non-verified data based on a calculation – score 3
iv) Documented estimate – score 4
v) Undocumented estimate – score 5

b) What is the status quo for the ownership and continuous support of data?

i) Hosts and owns – score 1
ii) Owns but does not host – score 2
iii) Hosts but does not own – score 3
iv) Hosts and owns partially – score 4
v) Does not host or own – score 5

c) Is the data regularly updated?

i) Regular updates – score 1
ii) Less frequent updates – score 2
iii) No updates – score 3

d) Is the data of deterministic nature, or are there statistically established confidence intervals stated for the data?

i) Confidence intervals developed considering parameter, scenario, and model uncertainty based on directly measured or calculated data – score 1
ii) Confidence intervals developed considering parameter, scenario, and/or model uncertainty based on an assumed probability distribution – score 2
iii) Deterministic value provided – score 3

Data collection methods

Data collection methods are assessed at the flow level, and they reflect the robustness of the sampling methods used (e.g., sample size) and the data collection period. In order to point at the specifics of the data collection methods, the pavement-specific pedigree matrix lists two questions within the data collection methods criterion, and the data quality assessment is carried out as follows:

a) Rounding representativeness to the nearest whole number, how representative of the market is the data?

i) ≥80% of the relevant market, over an adequate period – score 1
ii) 60-79% of the relevant market, over an adequate period OR representative data from ≥80% of the relevant market, over a shorter period – score 2
iii) 40-59% of the relevant market, over an adequate period OR representative data from 60-79% of the relevant market, over a shorter period – score 3
iv) <40% of the relevant market, over an adequate period OR representative data from 40-59% of the relevant market, over a shorter period – score 4
v) Unknown OR data from a small number of sites and from shorter periods – score 5

b) How compatible is the life cycle inventory data with the TRACI 2.1 impact assessment method from LCA Commons?

i) Enough to calculate all 9 mid-point indicators as per TRACI 2.1 – score 1
ii) Enough to calculate only 6 out of 9 mid-point indicators as per TRACI 2.1 – score 2
iii) Enough to calculate only 3 of 9 mid-point indicators as per TRACI 2.1– score 3
iv) Not compatible with TRACI 2.1 impact assessment method from LCA Commons – score 4

Time period of data

Time period is assessed at the flow level and is used for either: 1) assessing the age difference between the temporal data quality guidance and the age of the data, or 2) assessing the actual age of the data. In order to point at the specifics of time period, the pavement-specific pedigree matrix lists three questions within the time period criterion, and the data quality assessment is carried out as follows:

a) Does the data capture seasonal variations?
- i) All three (fall, spring, and summer) seasons are covered – score 1
- **ii) Only two out of three seasons are covered – score 2**
- iii) Only one season is covered – score 3
- iv) Not specified – score 4

b) How well is the time period of the data correlated with the data quality objective?
- i) Less than 3 years of difference – score 1
- **ii) Less than 6 years of difference – score 2**
- iii) Less than 10 years of difference – score 3
- iv) Less than 15 years of difference – score 4
- v) Age of data unknown or more than 15 years – score 5

c) How old was the data at the time of the data quality assessment?
- i) Less than 3 years old – score 1
- **ii) Less than 6 years old – score 2**
- iii) Less than 10 years old – score 3
- iv) Less than 12 years old – score 4
- v) Age of data unknown or more than 15 years – score 5

Question “b” is relevant for individual LCA studies with specific data quality objectives, whereas question “c” is relevant for assessing the data quality of background data without a specific data quality objective. As the scope of this roadmap is relevant to background data only, questions “a” and “c” are used to assess the data quality.

Geography of data

Geography is assessed at the flow level and is designed to capture differences in data quality related to differences in: 1) the resolution between the geography DQGs and the data used for modeling, and 2) the area of study. To point at the specifics of geography, the pavement-specific pedigree matrix lists two questions within the geography criterion, and the data quality assessment is carried out as follows:

a) How well is the geography of the data correlated with the data quality objective?
- i) Data from same resolution AND same area of study – score 1
- **ii) Within one level of resolution AND a related area of study – score 2**
- iii) Within two levels of resolution AND a related area of study – score 3
- iv) Outside of two levels of resolution BUT a related area of study – score 4
- v) From a different or unknown area of study – score 5

b) What is the regional granularity associated with the data?
- i) State level – score 1
- **ii) Country level – score 2**
- iii) Continental level – score 3
- iv) Global level – score 4
- v) Data granularity unknown – score 5

Question “a” is relevant for individual LCA studies with specific data quality objectives, whereas question “b” is relevant for assessing the data quality of background data without a specific data quality objective. As the scope of this roadmap is relevant to background data only, question “b” is used to assess the data quality.

Technology of data

Technology is assessed at the flow level and is designed to capture process design, operating conditions, material quality, and process scale. In order to point at the specifics of technology, the pavement-specific pedigree matrix lists two questions within the technology criterion, and the data quality assessment is carried out as follows:

a) How well is the technology of the data correlated with the data quality objective?

i) All technology categories are equivalent – score 1
ii) Three of the technology categories are equivalent – score 2
iii) Two of the technology categories are equivalent – score 3
iv) One of the technology categories is equivalent – score 4
v) None of the technology categories are equivalent – score 5

b) How well is the technology of the data described?

i) Specified – score 1
ii) Not specified – score 2

Question “a” is relevant for individual LCA studies with specific data quality objectives, whereas question “b” is relevant for assessing the data quality of background data without a specific data quality objective. As the scope of this roadmap is relevant to background data only, question “b” is used to assess the data quality.

Process level

Process level review enables assessment of the level of detail pertaining to a unit process; e.g., whether it is possible to obtain specific unit process information, or if only aggregated process information (combined processes to maintain confidentiality) is available.

Process review

Process review is assessed at the process level and is designed to evaluate the level of review a dataset has undergone at the unit process level. In order to point at the specifics of process review, the pavement-specific pedigree matrix lists one question within the process review criterion, and the data quality assessment is carried out as follows:

a) How well is the process reviewed?

i) Documented reviews by a minimum of two types of third-party reviewers – score 1
ii) Documented reviews by a minimum of two types of reviewers, with one being a third party – score 2
iii) Documented review by a third-party reviewer – score 3
iv) Documented review by an internal reviewer – score 4
v) No documented review – score 5

Process completeness

Process review is assessed at the process level and is designed to evaluate the level of review a dataset has undergone at the unit process level. In order to point at the specifics of process review, the pavement-specific pedigree matrix lists one question within the process review criterion, and the data quality assessment is carried out as follows:

b) With completeness rounded to the nearest whole value, how complete is the process?

i) >80% of determined flows within the process have been evaluated and given a value – score 1
ii) 60-79% of determined flows within the process have been evaluated and given a value – score 2
iii) 40-59% of determined flows within the process have been evaluated and given a value – score 3
iv) <40% of determined flows within the process have been evaluated and given a value – score 4
v) Process completeness not scored – score 5

Resource: *Recommended LCIs*

The following LCIs prescribed for fossil fuels, transportation, equipment, and electricity can be found at: https://www.lcacommons.gov/lca-collaboration/

Fossil fuels used in production

Diesel

Technologies relevant for the background data category 'Diesel'

- Diesel, combusted in industrial boiler
- Diesel, combusted in industrial equipment

Diesel meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
NETL (limited meta-data)	OpenLCA	Yes (based on NETL's CO2U Project)	Yes	No
USLCI	OpenLCA	Yes (based on non-verified data)	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Diesel from the USLCI's NREL is available on the LCA Commons collaboration server.

Gasoline

Technologies relevant for the entity 'Gasoline'

- Gasoline, combusted in industrial equipment

Gasoline meta-data on LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
NETL (no meta-data)	OpenLCA	Yes (based on NETL's CO2U Project)	Yes (non-verified)	No
GREET	GREET (EXCEL)	No (only GHG)	No	Yes
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Gasoline from the USLCI's NREL is available on the LCA Commons collaboration server.

Liquefied petroleum gas

Technologies relevant for the entity 'Liquefied petroleum gas'

- Liquefied petroleum gas, combusted in industrial boiler

Liquefied petroleum gas meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Liquefied petroleum gas from the USLCI's NREL is available on the LCA Commons collaboration server.

Residual fuel oil

Technologies relevant for the entity 'Residual fuel oil'

- Residual fuel oil, combusted in industrial boiler

Residual fuel oil meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Residual fuel oil from the USLCI's NREL is available on the LCA Commons collaboration server.

Coal

Technologies relevant for the entity 'Coal'

- Anthracite coal, combusted in industrial boiler
- Bituminous coal, combusted in industrial boiler
- Lignite coal, combusted in industrial boiler

Coal meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Coal from the USLCI's NREL is available on the LCA Commons collaboration server.

Recycled oil

This inventory was developed using the data in the consequential LCA by Geyer, Roland & Kuczenski, Brandon & Henderson, Ashley & Zink, Trevor (2013); and Life Cycle Assessment of Used Oil Management in California Pursuant to Senate Bill 546 (Lowenthal).

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
Geyer et al. (2013)	Presented in OpenLCA	Data source has emissions that have been converted to a TRACI compliant inventory.	Yes	Yes

Transportation

Diesel powered

Technologies relevant for the background data category 'Diesel'

- Transportation by barge, diesel powered
- Transportation by combination truck, diesel powered
- Transportation by ocean freighter, diesel powered
- Transportation by train, diesel powered
- Transport, refuse truck, diesel powered

Diesel meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
NETL (limited meta-data)	OpenLCA	Yes	Yes	No
USLCI	OpenLCA	Yes (Based on non-verified data)	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (Only GHG)	No	Yes

Diesel from the USLCI's NREL is available on the LCA Commons collaboration server.

Gasoline

Following technologies are relevant for the entity 'Gasoline'

- Transportation by combination truck, gasoline powered
- Transport, refuse truck, gasoline powered

Gasoline meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USLCI	OpenLCA	Yes	Yes (non-verified)	Yes (non-verified)
NETL (No meta-data)	OpenLCA	Yes (Based on NETL's CO2U Project)	Yes (non-verified)	No
GREET	GREET (EXCEL)	No (Only GHG)	No	Yes

Gasoline from the USLCI's NREL is available on the LCA Commons collaboration server.

Equipment

Equipment using diesel

Technologies relevant for the background data category 'Diesel'

- Operation of diesel equipment, industry average <19kW
- Operation of diesel equipment, industry average >19 kW and <56 kW
- Operation of diesel equipment, industry average >56 kW and <560 kW
- Operation of diesel equipment, industry average >560 kW and <900 kW
- Operation of diesel equipment, industry average >900 kW

Diesel meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USEPA-USLCI-GREET	OpenLCA	Yes	Yes	Yes

Diesel used in heavy construction equipment from USEPA is available on the LCA Commons collaboration server.

Equipment using natural gas

Technologies relevant for the background data category 'Natural gas'

- Operation of compressed natural gas equipment, industry average >19 kW and <56 kW
- Operation of compressed natural gas equipment, industry average >56 kW and <560 kW

Natural gas meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USEPA-USLCI-GREET	OpenLCA	Yes	Yes	Yes

Equipment using gasoline

Technologies relevant for the entity 'Gasoline'

- Operation of gasoline equipment, 2-stroke, industry average <19 kW
- Operation of gasoline equipment, 4-stroke, industry average <19 kW
- Operation of gasoline equipment, industry average <19 kW and >56 kW
- Operation of gasoline equipment, industry average >56 kW and <560 kW

Gasoline meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USEPA-USLCI-GREET	OpenLCA	Yes	Yes	Yes

Equipment using liquified petroleum gas

Technologies relevant for the entity 'Liquefied petroleum gas'

- Operation of liquefied petroleum gas equipment, industry average >19 kW and <56 kW
- Operation of liquefied petroleum gas equipment, industry average >56 kW and <560 kW

Liquefied petroleum gas meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
USEPA-USLCI-GREET	OpenLCA	Yes	Yes	Yes

Electricity

Technologies relevant for the background data category 'Electricity'

- Generation-based electricity from GREET and United States Lifecycle Inventories' (USLCI's) National Renewable Energy Laboratory (NREL)
- Consumption-based electricity from the National Energy Technology Laboratory (NETL)

The life cycle inventory data for consumption-based electricity is available from the National Energy Technology Laboratory (NETL), and the generation-based electricity is available from GREET as well as United States Lifecycle Inventories' (USLCI's) National Renewable Energy Laboratory (NREL). The project team has decided to choose consumption-based electricity (i.e., what electricity is consumed) from NETL over generation-based (i.e., what electricity is generated) as the default background data for electricity based on conversations with LCA Commons. However, this data is still not available in an OpenLCA-compatible format.

Electricity meta-data from LCA Commons stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
NETL	Grid-Mix Explorer (EXCEL)	Yes	Yes	No
USLCI	OpenLCA	Yes (based on non-verified data)	Yes (non-verified)	Yes (non-verified)
GREET	GREET (EXCEL)	No (only GHG)	No	Yes

Resource: *Example of proprietary dataset made public*

This example illustrates how system-level inventories can be made available by industries to help support downstream and associated products in the supply chain, while also protecting proprietary product information, thus creating more confidence in the results reported in the EPD. The Asphalt Institute has provided system level inventories for asphalt binder, which is used as an input for the asphalt mixture inventory and is publicly available through the Federal LCA Commons.

Asphalt binder

Technologies relevant for the entity 'Asphalt binder':

- Asphalt binder, no additives, consumption mix, at terminal, from crude oil
  (Source: http://www.asphaltinstitute.org/engineering/lca-study-on-asphalt-binders/)
- Asphalt binder, 0.5% polyphosphoric acid (PPA), consumption mix, at terminal, from crude oil
  (Source: http://www.asphaltinstitute.org/engineering/lca-study-on-asphalt-binders/)
- Asphalt binder, 3.5% styrene-butadiene-styrene (SBS), consumption mix, at terminal, from crude oil
  (Source: http://www.asphaltinstitute.org/engineering/lca-study-on-asphalt-binders/)
- Asphalt binder, 8% ground rubber tire (GRT), consumption mix, at terminal, from crude oil
  (Source: http://www.asphaltinstitute.org/engineering/lca-study-on-asphalt-binders/)

Asphalt binder meta-data from stakeholders

Sources	Formats	TRACI mid-point indicators compliance	OpenLCA LCI	Energy
Asphalt Institute	OpenLCA	Yes, compatible with TRACI 2.1 from LCA Commons	Yes	No

Asphalt institute's LCA team worked with NREL to make their life cycle inventory compatible with the LCA Commons TRACI 2.1 impact assessment method, and the dataset has recently been added to the LCA Commons collaboration server.

References

Bhat, Chaitanya Ganesh, Life Cycle Information Models with Parameter Uncertainty Analysis to Facilitate the Use of Life-Cycle Assessment Outcomes in Pavement Design Decision-Making, Open Access Dissertation, Michigan Technological University, 2020.

Edelen, A.; Ingwersen, W. Guidance on Data Quality Assessment for Life Cycle Inventory Data; Life Cycle Assessment Paper Center, National Risk Management Paper Laboratory, United States Environmental Protection Agency, 2016. Available here.

Ingwersen, W. and Stevenson, M., 2012. Can we compare the environmental performance of this product to that one? An update on the development of product category rules and future challenges toward alignment. Journal of Cleaner Production, 24, pp.102-108.

Sphera (Thinkstep) (2019) Life Cycle Assessment of Asphalt Binder, Asphalt Institute. Willis, J.R. (2015) Effect of Tool Choice on LCA Results. Presented December 8, 2015, at FHWA Sustainable Pavements Technical Working Group Meeting #10, Chicago, Illinois.