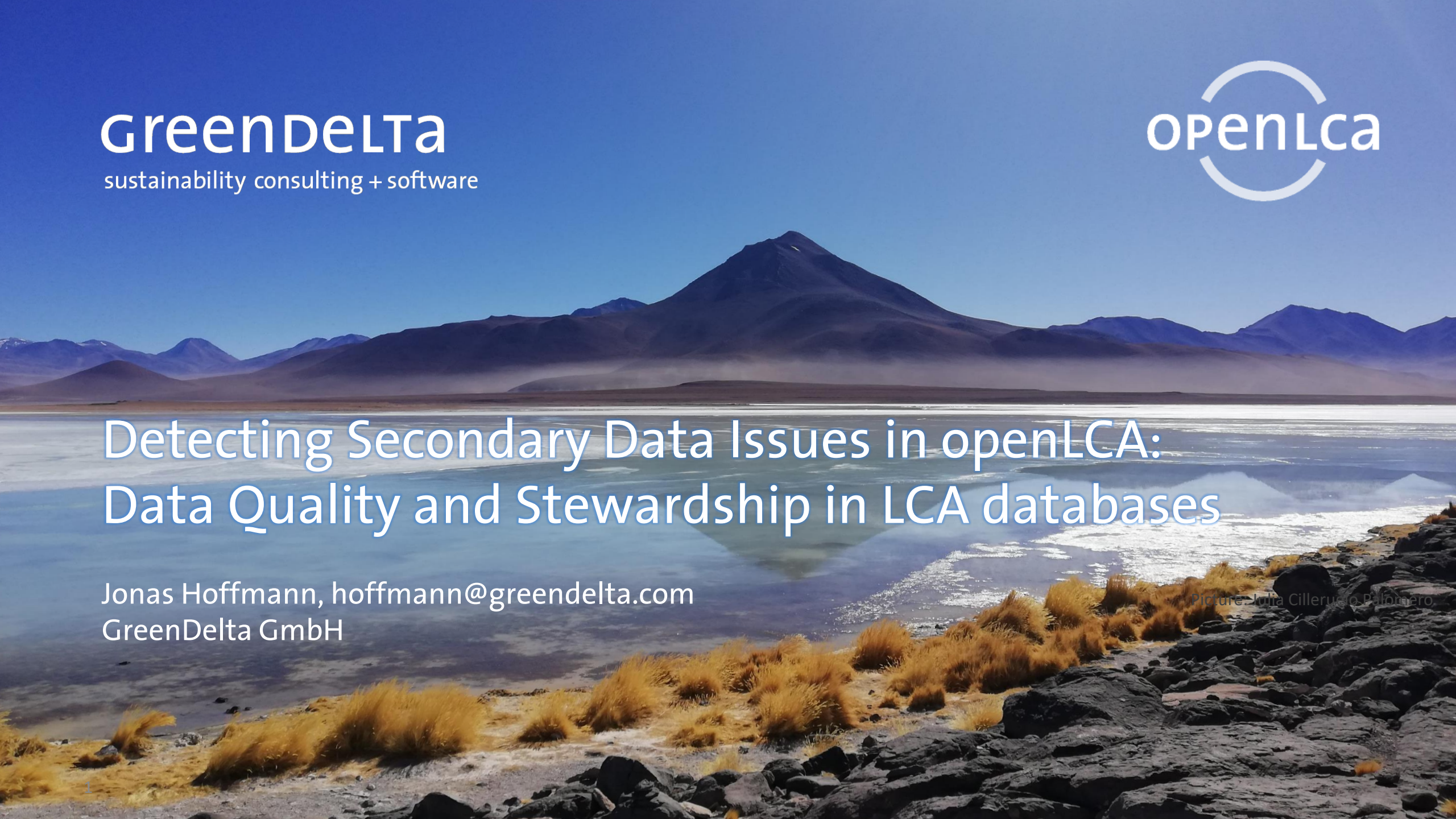


GreenDelta
sustainability consulting + software

openLCA

# Detecting Secondary Data Issues in openLCA: Data Quality and Stewardship in LCA databases

Jonas Hoffmann, hoffmann@greendelta.com
GreenDelta GmbH

Picture: Julia Cillero Palomero

# Motivation

Large differences between LCA background databases do not happen randomly. They usually have identifiable causes. A good starting point is to carefully review the dataset documentation. In particular, look at:

a) the technical description of the dataset and whether the name accurately reflects its content,
b) the types of sources used
c) the age of the underlying data, not to be confused with the dataset's release date,
d) how professionally the data was compiled, quality assurance and any external review

If anything remains unclear, it is better to contact the data provider directly rather than speculate or make assumptions. There is always an explanation, and it can usually be clarified. Professional data providers understand their datasets, sources, and system boundaries very well.

# Motivation

| Framework | Use of secondary data | Reviewer/Verifier |
|---|---|---|
| ISO 14040 / 14044 | • Must meet defined **data quality requirements** (temporal, geographical, technological representativeness, completeness, consistency)<br>• Limitations must be documented | • Checks whether the **data quality requirements** were defined consistently with the goal and scope the selected secondary data are fit for purpose<br>• whether data sources, assumptions, limitations, and cut-offs are transparently documented |
| EPD International (GPI) | • Data **must be recent (<5 years)**, PCR compliant, and sourced from LCI databases<br>• data shall be **checked for plausibility** (e.g., by mass or energy balance, or by comparisons with other relevant sources)<br>• Overuse of generic data is not acceptable | • Checks that **secondary data are relevant** and consistent with the applicable PCR/GPI<br>• Their use is transparent, plausible and not excessive where primary data are required<br>• **Sample checks** should be carried out for unit processes that have a significant influence<br>• **Random sample** of unit processes/modules |
| Product Environmental Footprint (PEF) | • Use of **approved EF compliant secondary datasets** is mandatory (see PEFCR)<br>• Substitution or deviation of datasets is limited and must be justified (**maximum of 10% of the single score** may derived from ILCD-EL compliant datasets) | • **Reviewer validates 80% of the most-relevant processes** and the datasets used to model these processes |

## What to expect

- **With upcoming regulatory pressure, secondary data becomes more relevant**

- How do different database work, what is their data quality approach, how to catch issues and how data inconsistencies were managed:
  - **EF 3.1 database**
    - Articulated lorry transport, Total weight, <7.5 t, mix Euro 0-5
  - **CarbonMinds**
    - Palm oil supply chain and o-toluenediamine production
  - **Ecoinvent**
    - Production of fluorinated gases affecting the supply chain

carbonminds

ecoinvent

- Further, best-practice in **data stewardship** will be debated

# EF 3.1 database

- Covers all sectors (from transport, to electricity, to chemicals etc.)
- The EF 3.1 database should be used exclusive for PEF studies
  - Data has been provided by various providers building onto each other
  - Core datasets are derived from Sphera further from CEPE, ESIG, ecoinvent, Blonk
  - Data is uploaded on individual nodes (soda4LCA) an depending on the data provider, various aggregation levels are found (fully, level-1 disaggregated)

| | | | |
|---|---|---|---|
| Small Data Providers Database | **Node operated by the European Commission for small data providers (less than 10 process datasets per provider allowed)**<br><br>(EF Pilot phase data)<br><br>(EF 3.1 Hydrogen data) | European Commission | https://eplca.jrc.ec.europa.eu/EF-SDP/ |
| Sphera (former Thinkstep) | **Core datasets** official ETPE (includes Energy, Transport, Packaging, End-of-life) (EF 2.0) (tendered, EF pilot phase)<br>**Core datasets** official ETPE part 1 (EF 3.1) (updated from EF pilot phase)<br>**Core datasets** official ETPE part 2 (includes non-packaging plastics, electric and electronics, metals and minerals) (EF 3.1) (tendered, EF transition phase) | Sphera | http://lcdn.thinkstep.com/ |

  - Data quality is provided on process level (will be detailed)

# EF 3.1 database:

Articulated lorry transport, Euro 0, Total weight 7,5-12 t, consumption ...

**Inputs/Outputs - Articulated lorry transport, Euro 0, Total weight 7,5-12 t, consumption mix, to consumer, diesel driven, 1980s, cargo, 7,5 t - 12t gross weight / 5t payload capacity - EU+EFTA+UK**

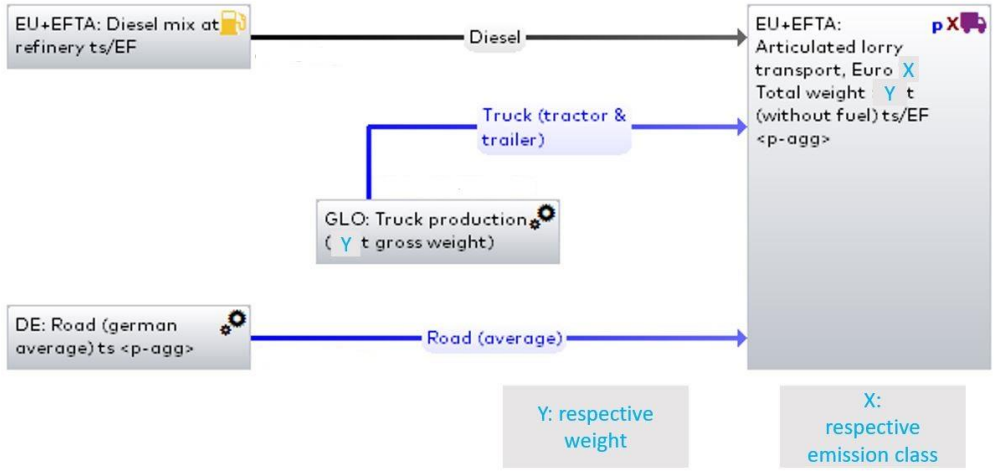
**Inputs**

| Flow | Category | Amount | Unit | Costs/R... | Uncertai... | Avoided... | Provider | Data qu... | Location | Descript... |
|---|---|---|---|---|---|---|---|---|---|---|
| Water (rain water) | Resources/Resourc... | 3.98588 | kg | | none | | | | | |
| crude oil | Resources/Resourc... | 2.48987 | MJ | | none | | | | | |
| natural gas | Resources/Resourc... | 0.22273 | MJ | | none | | | | | |
| primary energy f... | Resources/Resourc... | 0.15895 | MJ | | none | | | | | |
| natural aggregate | Resources/Resourc... | 0.10153 | kg | | none | | | | | |
| air | Resources/Resourc... | 0.05908 | kg | | none | | | | | |
| sea water | Resources/Resourc... | 0.04463 | kg | | none | | | | | |
| calcium carbonate | Resources/Resourc... | 0.03882 | kg | | none | | | | | |
| inert rock | Resources/Resourc... | 0.03826 | kg | | none | | | | | |

**Outputs**

| Flow | Category | Amount | Unit | Costs/R... | Uncertai... | Avoided... | Provider | Data qu... | Location | Descript... |
|---|---|---|---|---|---|---|---|---|---|---|
| **Transport** | **Transport services** | **1000.0...** | **kgkm** | | **none** | | | | | |
| High radioactive... | Wastes/Radioactiv... | 9.72451... | kg | | none | | | | | |
| Low radioactive ... | Wastes/Radioactiv... | 1.29071... | kg | | none | | | | | |
| Medium radioac... | Wastes/Radioactiv... | 6.08431... | kg | | none | | | | | |
| Radioactive tailin... | Wastes/Radioactiv... | 6.68834... | kg | | none | | | | | |
| (2,4-dichlorophe... | Emissions/Emissio... | 2.89951... | kg | | none | | | | | |
| (2,4-dichlorophe... | Emissions/Emissio... | 7.61422... | kg | | none | | | | | |
| (R)-2-(4-chloro-... | Emissions/Emissio... | 6.90109... | kg | | none | | | | | |
| 1,1'-dimethyl-4,... | Emissions/Emissio... | 3.49101... | kg | | none | | | | | |

General information | Inputs/Outputs | Documentation | Parameters | Allocation | Social aspects | Direct impacts | Additional properties




It is very hard to understand the scope and inputs of the dataset based in the description and provided details.

How about Data Quality?



GreenDelta

# EF 3.1 database: Data quality

- Data quality per dataset and review:

| Criterion | Dataset Value | Review Statement |
|---|---|---|
| Precision | Good (2.0) | Good (2.0) |
| Geographical representativeness | Very good (1.0) | Very good (1.0) |
| Time representativeness | Very good (1.0) | Very good (1.0) |
| Technological representativeness | Very good (1.0) | Very good (1.0) |
| Overall quality | Very good (1.3) | Very good (1.3) |

- Reality:

Start date 01.01.2012 , **diesel driven, 1980s, cargo,**

“Annual average. The DQR of the dataset reflects the quality of the data at the time of release. **The user of the dataset should revise the DQR at the moment of application.**”




GreenDelta

# EF 3.1 database: Aggregation of data and data stewardship

Inputs/Outputs: Articulated lorry transport, Total weight 7.5-12 t, mix Euro 0-5, consumption mix, to consumer, diesel driven, Euro 0 - 5 mix, cargo, 7,5 - 12t gross weight / 5t payload capacity - EU+EFTA+UK

**Inputs**

| Flow | Category | Amount | Unit | Costs/Reve... | Uncertainty | Avoided w... | Provider | Data qualit... | Location | Description |
|---|---|---|---|---|---|---|---|---|---|---|
| Electricity; consumpti... | Energy carriers and tech... | 0.00063 | MJ | | none | | | | | |
| Thermal Energy (EU) | Energy carriers and tech... | 0.00294 | MJ | | none | | | | | |
| Steel scrap (St) | End-of-life treatment | 5.02320E-5 | kg | | none | ☐ | | | | |
| agriculture | Land use/Land occupati... | 2.75956E-14 | m2*a | | none | | | | | |
| air | Resources from air/Ren... | 0.05412 | kg | | none | | | | | |
| aluminium | Resources from ground... | 5.63574E-11 | kg | | none | | | | | |
| antimonite | Resources from ground... | 8.85510E-21 | kg | | none | | | | | |
| antimony | Resources from ground... | 5.56416E-9 | kg | | none | | | | | |
| arable | Land use/Land occupati... | 0.01190 | m2*a | | none | | | | | |
| arable, irrigated | Land use/Land occupati... | 2.15996E-14 | m2*a | | none | | | | | |
| arable, irrigated, inten... | Land use/Land occupati... | 2.46120E-8 | m2*a | | none | | | | RO | |
| arable, irrigated, inten... | Land use/Land occupati... | 3.85053E-6 | m2*a | | none | | | | ES | |
| arable, irrigated, inten... | Land use/Land occupati... | 1.34249E-6 | m2*a | | none | | | | NL | |
| arable, irrigated, inten... | Land use/Land occupati... | 0.00015 | m2*a | | none | | | | DE | |
| arable, irrigated, inten... | Land use/Land occupati... | 4.31623E-6 | m2*a | | none | | | | BR | |
| arable, irrigated, inten... | Land use/Land occupati... | 2.04033E-5 | m2*a | | none | | | | GB | |
| arable, irrigated, inten... | Land use/Land occupati... | 2.87255E-6 | m2*a | | none | | | | PL | |
| arable, irrigated, inten... | Land use/Land occupati... | 0.01080 | m2*a | | none | | | | | |
| arable, irrigated, inten... | Land use/Land occupati... | 8.30963E-9 | m2*a | | none | | | | IT | |
| arable, irrigated, inten... | Land use/Land occupati... | 2.78289E-5 | m2*a | | none | | | | FR | |
| arable, irrigated, inten... | Land use/Land occupati... | 0.00219 | m2*a | | none | | | | US | |
| arable, non-irrigated | Land use/Land occupati... | 1.29185E-16 | m2*a | | none | | | | | |
| arable, non-irrigated, ... | Land use/Land occupati... | 5.05767E-15 | m2*a | | none | | | | | |
| arable, non-irrigated, ... | Land use/Land occupati... | 5.23910E-5 | m2*a | | none | | | | DE | |
| arable, non-irrigated, ... | Land use/Land occupati... | 6.96772E-7 | m2*a | | none | | | | | |
| arable, non-irrigated, ... | Land use/Land occupati... | 1.87509E-7 | m2*a | | none | | | | BR | |
| argon | Resources from air/Ren... | 6.57615E-13 | kg | | none | | | | | |
| arsenic | Resources from ground... | 5.63630E-15 | kg | | none | | | | | |
| asbestos (white) | Resources from ground... | 3.66530E-15 | kg | | none | | | | | |
| barium sulfate | Resources from ground... | 1.15684E-12 | kg | | none | | | | | |

**Outputs**

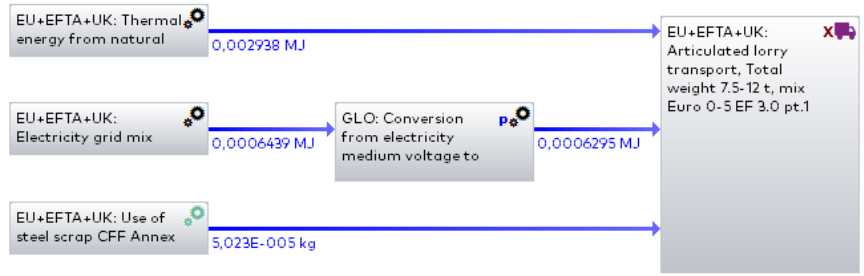
| Flow | Category | Amount | Unit | Costs/Reve... | Uncertainty | Avoided pr... | Provider | Data qualit... | Location | Description |
|---|---|---|---|---|---|---|---|---|---|---|
| **Transporting capacity** | **Transport services** | **1000.00000** | **kgkm** | | **none** | | | | | |

The fix was:

**Option 1**: Accept the imprecision. Impacts affected by ~1% for most categoriesTransport typically contributes <<10% of total impactOverall influence on PEF results likely <<0.1%

**Option 2**: Select an alternative

**Option 3**: Manually connect open input flows. User links flows to the 4 required processes




Meanwhile, the dataset has been reuploaded, flows have been deleted and changed without communicating this.

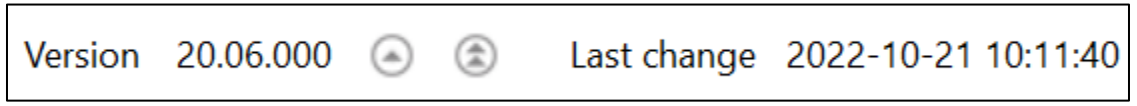
**Perform consistency checks (LCI)**



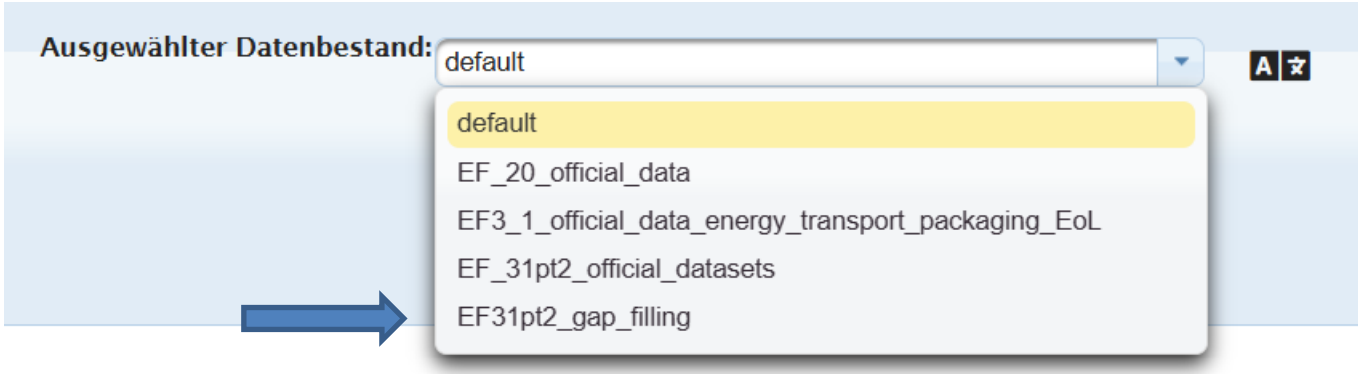
GreenDelta

## EF 3.1 database: Data Stewardship

- Changes of datasets should be communicated in a transparent fashion!
- The ILCD format allows versioning
- During the import of changed datasets openLCA will automatically overwrite the older version with a newer version!
- Every time you work with openLCA and change/save a dataset, it will update as stated in the ILCD format




- Date: Date when the data set was revised for the last time, typically manually set.
- Version: Version number of data set. First two digits refer to major updates, the second two digits to minor revisions and error corrections etc.




https://eplca.jrc.ec.europa.eu/LCDN/downloads/ILCD_Format_1.1_Documentation/ILCD_ProcessDataSet.html



GreenDelta

## CarbonMinds database

- Focuses on chemicals, polymers and other materials
- Types of processes (consumption/production mix, technology (core, simple, ext.)
- Data is fully aggregated but the description gives sufficient details:

**Technology**

| | |
|---|---|
| Description | The reaction of m-cresol and methyl isocyanate produces the main product (3-methylphenyl) n-methylcarbamate. For this purpose, the process consumes hydrochloric acid, ammonia, caustic soda (50%), calcium oxide, m-cresol, methyl isocyanate, titanium dioxide, and chromium oxide. Utilities consumed comprise thermal energy, electricity, unspecific inorganic chemical, cement, and disposal of cement. Utility demands in form of thermal energy and electricity are estimated according to the chemical park in Gendorf. For each kg of chemical product, 1.2 MJ electricity and 2 MJ thermal energy are required. A conversion rate of 95% is assumed. Process wastes are treated in municipal waste incinerators.<br>Background modeling: |

- DQR is according to TfS and often between 1.2-2.0 (good to fair):

**Review methods**

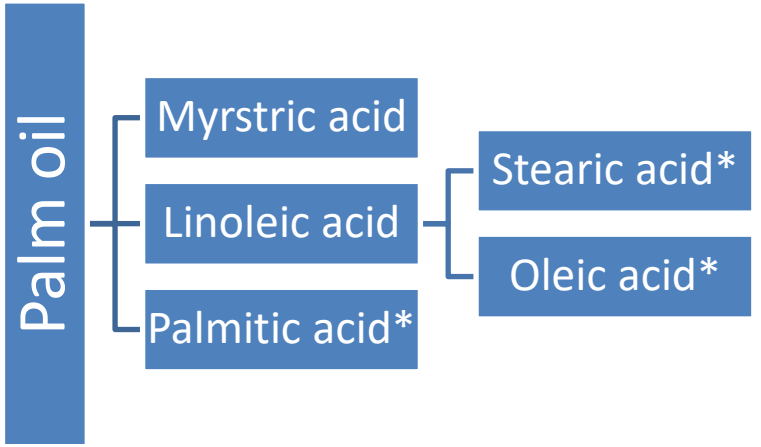
| Scope | Method |
|---|---|
| Documentation | Compliance with ISO 14040 to 14044 |
| LCI results or Partly terminated system | Compliance with ISO 14040 to 14044 |
| | Cross-check with other source |
| Life cycle inventory methods | Compliance with ISO 14040 to 14044 |
| | Expert judgement |
| Raw data | Cross-check with other source |
| | Validation of data sources |

**Quality assessment**

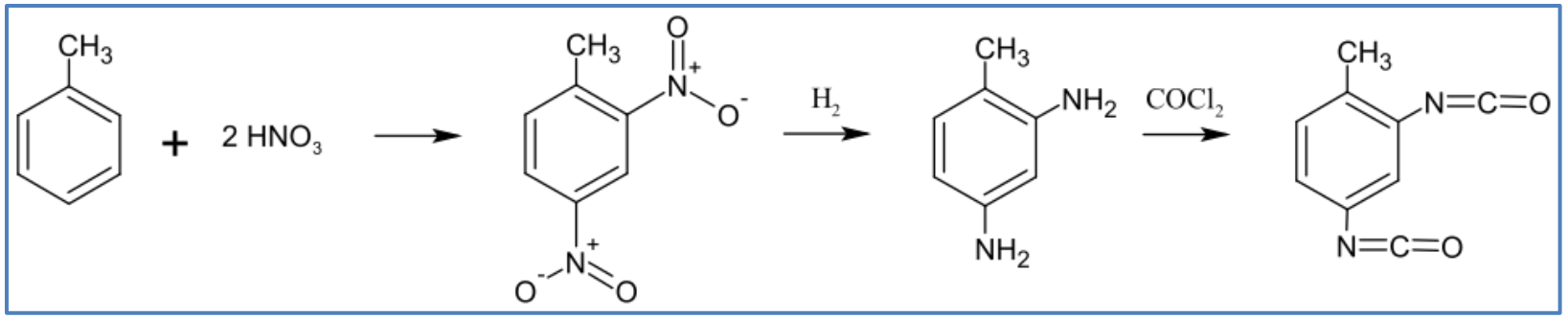
| Aspect | Very good | Good | Fair | Poor | Very poor | Not evaluated ... | Not applicable |
|---|---|---|---|---|---|---|---|
| Completeness | ☐ | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ |
| Geographical representativeness | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Methodological appropriateness an... | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

## CarbonMinds database: Aggregation of data and data stewardship

- Very hard due to aggregation of datasets
- For various fatty acids, the documentation is insufficient
- The product carbon footprints were analysed
  - Share on the impacts on two products and other were cut-off?
  - It is **unclear how to process was modelled**




- For o-toluenediamine, the production route suggested 'TDI production from toluene (CO from natural gas)' (technology-specific (extension layer))




- The o-toluenediamine can be potentially produced as a by-product, was it economically allocated? (Price is very high if it would be derived from TDI)

→ Inconsistencies were reported to CM

## ecoinvent database

- The data sets are available as **unit processes**, which increases transparency and allows users to trace impacts along the supply chain
- The database covers a broad range of sectors and geographies, mainly representing average production rather than company-specific
- Relicts from **ecoinvent version 2** can still matter in practice because version 2 historically influenced many datasets and workflows
- Processes typically represent **average market or production situations** for a geography, including global datasets where regional data is limited.
- Data quality is commonly handled using the **pedigree-matrix approach** on **flow level**

**Time**

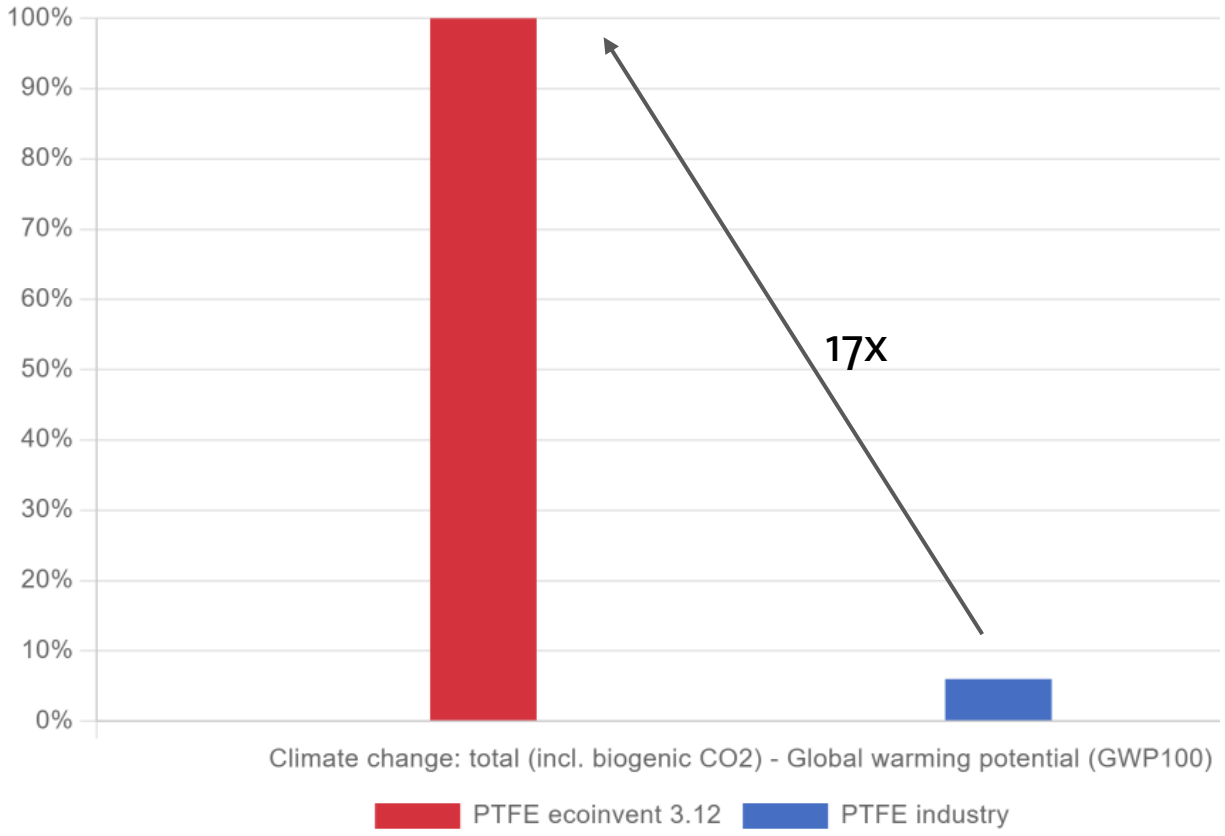
Start date 01.01.1999 End date 31.12.2024

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Reliability | Verified data based on measurements | Verified data partly based on assumptions or non-verified data based on | Non-verified data partly based on qualified estimates | Qualified estimate (e.g. by industrial expert) | Non-qualified estimates |
| Completeness | Representative data from all sites relevant for the market considered, over and adequate period to even out normal fluctuations | Representative data from > 50% of the sites relevant for the market considered, over an adequate period to even out normal fluctuations | Representative data from only some sites (<< 50%) relevant for the market considered or > 50% of sites but from shorter periods | Representative data from only one site relevant for the market considered or some sites but from shorter periods | Representativeness unknown or data from a small number of sites and from shorter periods |
| Temporal correlation | Less than 3 years of difference to the time period of the data set | Less than 6 years of difference to the time period of the data set | Less than 10 years of difference to the time period of the data set | Less than 15 years of difference to the time period of the data set | Age of data unknown or more than 15 years of difference to the time period of the data set |
| Geographical correlation | Data from area under study | Average data from larger area in which the area under study is included | Data from area with similar production conditions | Data from area with slightly similar production conditions | Data from unknown or distinctly different area (North America instead of Middle East, OECD-Europe instead of Russia) |
| Further technological correlation | Data from enterprises, processes and materials under study | Data from processes and materials under study (i.e. identical technology) but from different enterprises | Data from processes and materials under study but from different technology | Data on related processes or materials | Data on related processes on laboratory scale or from different technology |

| Flow | Data quality e... | Description |
|---|---|---|
| heat, district or i... | (5; 3; 5; 5; 3) | Rough estimation |
| chlorodifluorom... | (4; 3; 5; 5; 1) | Stoichiometric calculation, 95% yield |
| chemical factory,... | (2; 3; 3; 2; 3) | Calculated based on literature data p |
| nitrogen, liquid | (2; 3; 3; 2; 3) | Calculated based on literature data p |

## ecoinvent database; the case of PTFE production

- Read the documentation carefully and perform sense checks (PCF)




The production of fluorochemicals and PTFE monomers can be summarized with the following chemical reactions: $CaF_2 + H_2SO_4 \rightarrow CaSO_4 + 2HF$ (1);$CH_4 + 3Cl_2 \rightarrow CHCl_3 + 3HCl$ (2);$CHCl_3 + 2HF \rightarrow CHClF_2 + 2HCl$ (3);$2\ CHClF_2 + heat \rightarrow CF_2=CF_2 + 2\ HCl$ (4); A large number of other by-products and emissions is formed in the processes **(benzene, dichloromethane, ethylene oxide, formaldehyde, R134a, and vinyl chloride)** and small amounts of the highly toxic perfluoroisobutylene $CF_2=C(CF_3)_2$. The by-products in the production of monomers can harm the processes of polymerisation. Because of this, the refinement of the production of monomers has to be very narrow. This makes the process complex and it contributes to a high cost for the PTFE-laminates. (Cedergren et al. 2001).;References:;Althaus H.-J., Chudacoff M., Hischier R., Jungbluth N., Osses M. and Primas A. (2007) Life Cycle Inventories of Chemicals. Final report ecoinvent data v2.0 No. 8. Swiss Centre for Life Cycle Inventories.

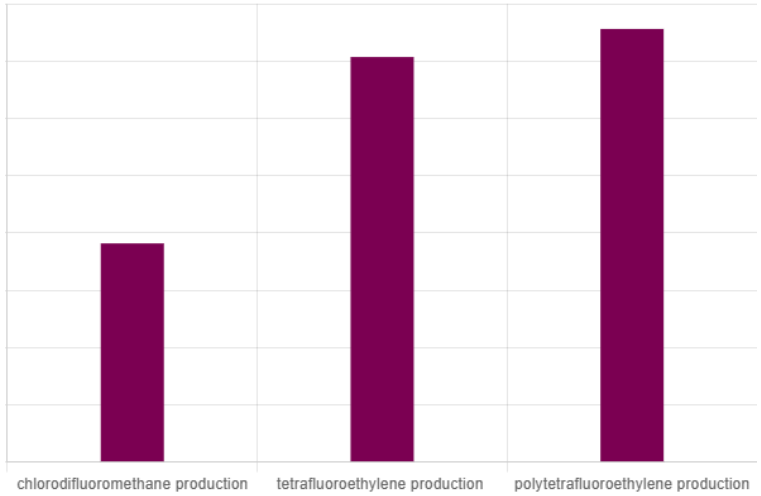
Aggregation of LCI data at this stage would hinder further investigation



GreenDelta

# ecoinvent database; legacy issues and data stewardship

- Detailed analysis:




Outputs

| Flow | Category | Description |
|---|---|---|
| Chlorodifluoromethane | ../Emission to air/high population density | Range 0.38-5% of product, estimation for European plant. |
| Chloroform | ../Emission to air/high population density | Literature 1994 |
| Dichlorodifluoromethane | ../Emission to air/high population density | Literature 1994 |
| Dichlorofluoromethane | ../Emission to air/high population density | Data for European plant. |
| Hydrochloric acid | ../Emission to air/high population density | Calculation in literature |
| Hydrogen fluoride | ../Emission to air/high population density | Calculation in literature |
| Trichlorofluoromethane | ../Emission to air/high population density | Literature 1994 |
| Trifluoromethane | ../Emission to air/high population density | Range 0.35-1.5% of product, estimation for European plan |
| Water | ../Emission to air/unspecified | Calculated value based on literature values and expert opi |

Outputs

| Flow | Category | Description |
|---|---|---|
| municipal solid waste | E:Water supply; sewerage, waste manage... | Literature, Ros 1994, manufacturing wa... |
| refinery sludge | E:Water supply; sewerage, waste manage... | Literature, Ros 1994, chemical waste an... |
| refinery sludge | E:Water supply; sewerage, waste manage... | Literature, Ros 1994, chemical waste an... |
| 1,1,1,2-Tetrafluoroethane | ../Emission to air/high population density | Estimation. 0.02% of the input of tetrafl... |
| Chlorodifluoromethane | ../Emission to air/high population density | Unspecified HCFC. Estimation based on... |
| Dichlorodifluoromethane | ../Emission to air/high population density | Estimation based on literature and expe... |
| Methanol | ../Emission to air/high population density | Literature |
| Nitrogen | ../Emission to air/unspecified | Calculated value. The amount of nitrog... |
| Trichlorofluoromethane | ../Emission to air/high population density | Estimation based on literature and expe... |

**IPCC Global Warming Potential (GWP) values relative to $CO_2$**

| Common chemical name or industrial designation | Chemical formula | GWP values for 100-year time horizon | | |
|---|---|---|---|---|
| | | Fourth Assessment Report (AR4) | Fifth Assessment Report (AR5) | Sixth Assessment Report (AR6) |
| **Major Greenhouse Gases** | | | | |
| Carbon dioxide | $CO_2$ | 1 | 1 | 1 |
| HFC-134a | $CH_2FCF_3$ | 1,430 | 1,300 | 1,530 |
| PFC-1114 | $CF_2=CF_2$ | | <1 | 0.004 |

$\tau(CF_2=CF_2) = 1\ d;\ \tau(CFH_2CF_3) = 14\ y$ (IPCC AR6)

**Due to the transparent unit processes**, the assumptions and one error was be found -> **ecoinvent enstated a working group on this**



GreenDelta

**Conclusions:**

- **Data quality is a shared responsibility**:

→ Practitioners must review datasets, but providers must ensure transparent stewardship

- **Aggregation reduces transparency**:

→ Limits understanding and often requires additional clarification from providers

- **Documentation alone is not enough (**mismatch data and documentation):

→ Always perform consistency checks (e.g., LCI, PCF)

- **Data stewardship is relevant**:

→ in only 1 out of 3 cases followed up or transparently described changes

- **Versioning and transparency remain critical**

→ Changes must be clearly communicated and traceable (ticketing)

- **See Validation Chapter: https://greendelta.github.io/openLCA2-manual/validation.html**

GREENDELTA
sustainability consulting + software

OPENLCA

# Thank you!

Picture: Julia Cilleruelo Palomero

GreenDelta GmbH
Alt-Moabit 130, 10557 Berlin
www.greendelta.com