GreenDelta

sustainability consulting + software

Introducing the LCA data machine

Andreas Ciroth, Michael Srocka

GreenDelta GmbH

SETAC Brussels, May 8, 2017

Introducing the LCA data machine

1. A motivation
2. Principles of the LCA data machine
3. Some examples
4. Status and outlook

1 A motivation

A motivation for the LCA data machine

(this is probably the easiest part of the entire presentation)

- LCA studies are data intense, modern studies integrate life cycles that cover > 10,000 individual process data sets
- Data collection and modelling is done manually, integrating expert judgement and many other sources, which is a lot of effort, error-prone, and difficult for quality assurance

A motivation for the LCA data machine

Data collection and modelling is a bottleneck for LCA

- Missing information
- Use of outdated information
- Use of incorrect information without being aware

→ Is there not another way to obtain LCA data?

The LCA data machine (LCADM)

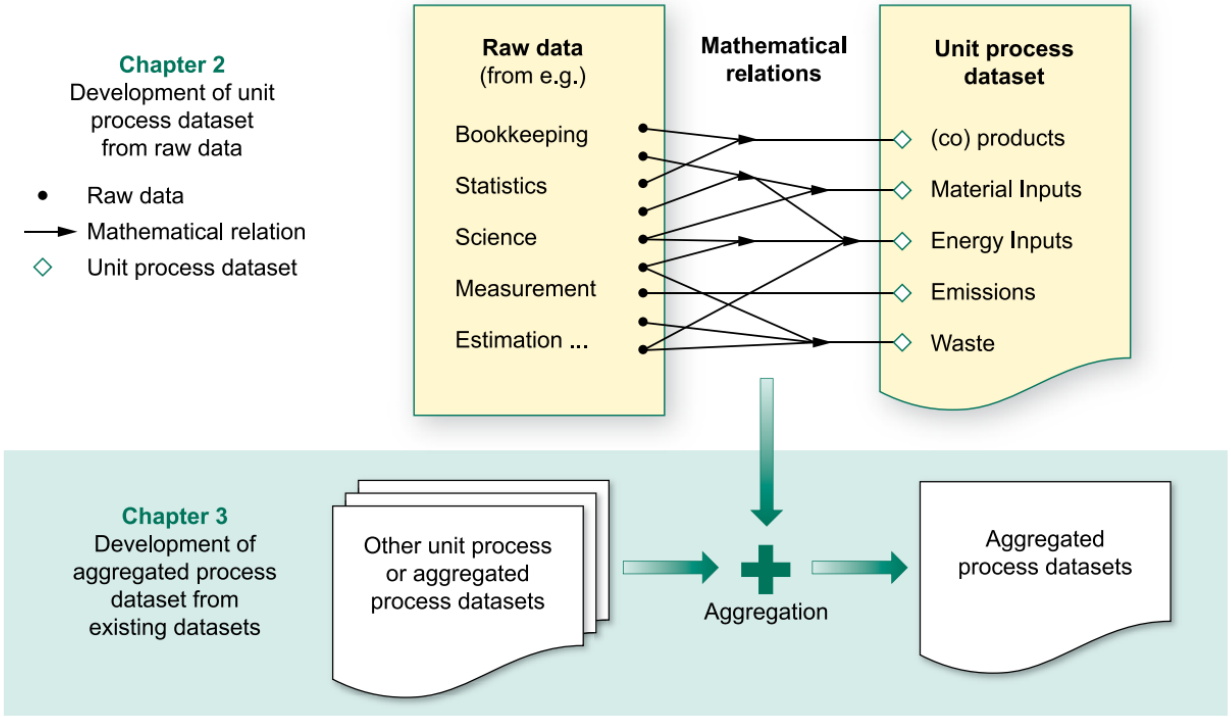
=

- A system to create data sets for LCA, automatically and fast, demand-driven, with controlled fitness for purpose
- Currently developed at GreenDelta
- German research project
  "Entwicklung eines autodidaktischen Data Mining Algorithmus mit hierarchischen temporalen Speicher (HTM) zur Interpolierung des [...] Footprints von Produkten", 2016-2018

2 The data machine, principles

The LCA data machine, principles (1/4)

- LCA datasets are created from raw data, various sources are combined, mathematical relations applied



Wang Ciroth et al. 2011, p 54

The LCA data machine, principles (2/4)

- Shannon entropy is used to assess **level of detail, information content** for processes, in terms of absolute information content H(A) and information gain

$$H(\boldsymbol{A}) = -\sum_{i=1}^{n} p_i * log_2 p_i$$

$$H(p|q) = -\sum_{i=1}^{n} p_i * log_2 \frac{p_i}{q_i}$$

The LCA data machine, principles (2/4)

- Shannon entropy is used to assess level of detail for processes, in terms of absolute information content H(A) and information gain

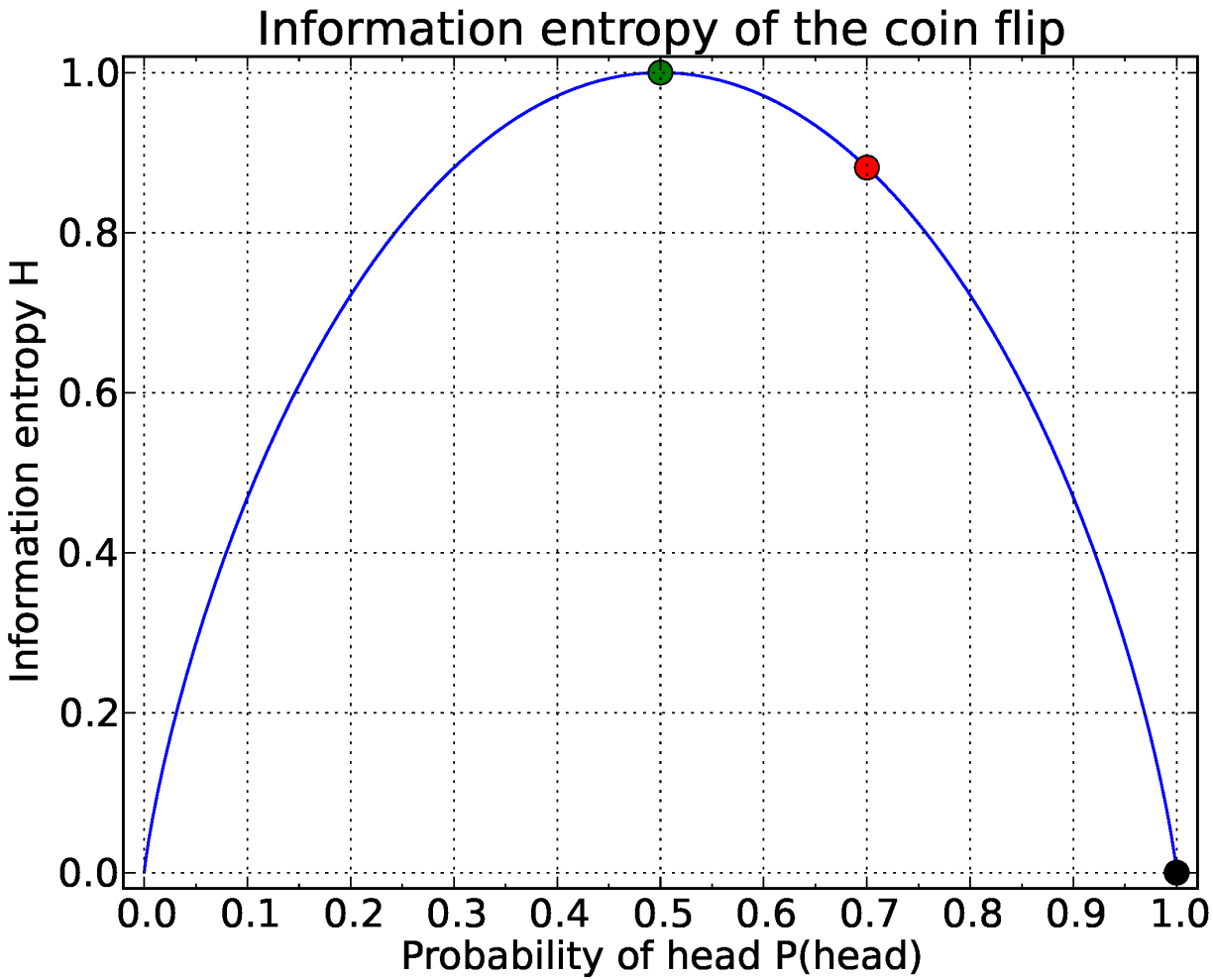
$$H(\boldsymbol{A}) = -\sum_{i=1}^{n} p_i * log_2 p_i$$

Information content

$$H(p|q) = -\sum_{i=1}^{n} p_i * log_2 \frac{p_i}{q_i}$$

Information gain: p in addition to q

Shannon entropy, classic example: coin flip



Wellmann, J.F.: Information Theory for Correlation Analysis and Estimation of Uncertainty Reduction in Maps and Models, *Entropy* **2013**, *15*(4), 1464-1485

The LCA data machine, principles (3/4)

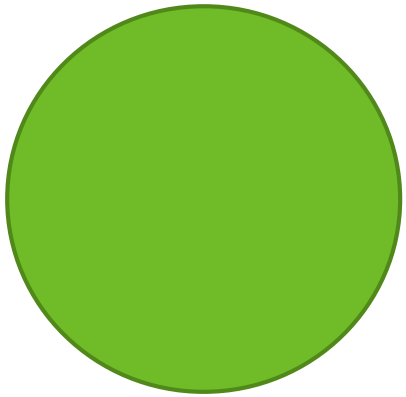
- Content conformance / **data quality / fitness for purpose** is used to measure the "quality" of the information

  (i.e. how well does the product, time, geography, modelling aspects fit to what is needed)

- Can contradict Shannon entropy

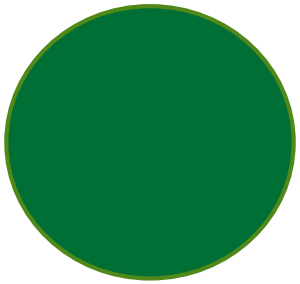
Shannon entropy vs fitness for purpose

- Broader information, less specific, in scope (i.e. fit for purpose)



Data set older than 2016

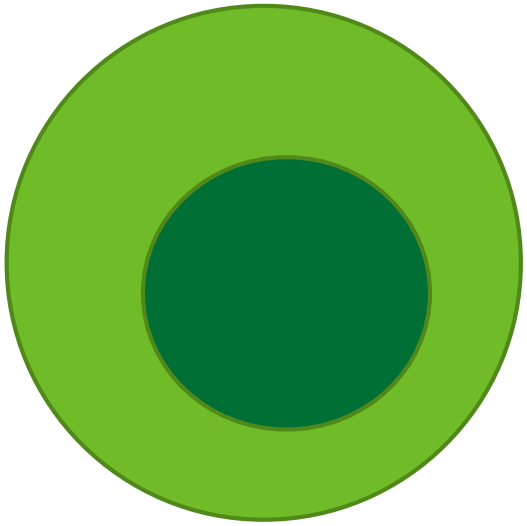
- More specific information



Data set from 2005

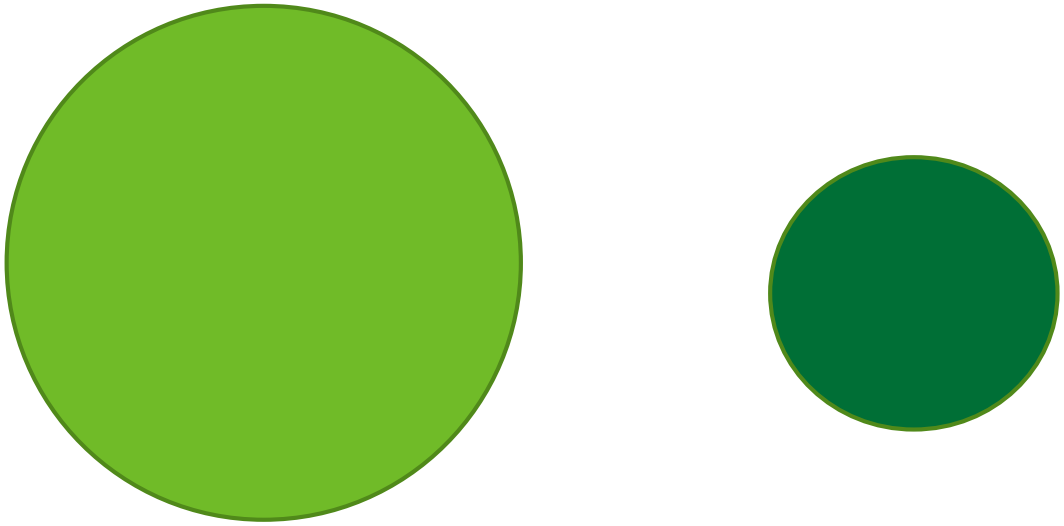
Shannon entropy vs fitness for purpose

- ***Case 1***: broader information in scope, more detailed information also in scope
  Shannon entropy ↘, fitness for purpose ↗, improvement



Shannon entropy vs fitness for purpose

- ***Case 2***: broader information in scope, more detailed information not in scope
  Shannon entropy ↘, fitness for purpose ↓, degradation



The LCA data machine, principles (4/4)
Patterns and rules defined

Process types (12 different so far:

fuel incineration,
cargo transport,
agriculture – husbandry

....)

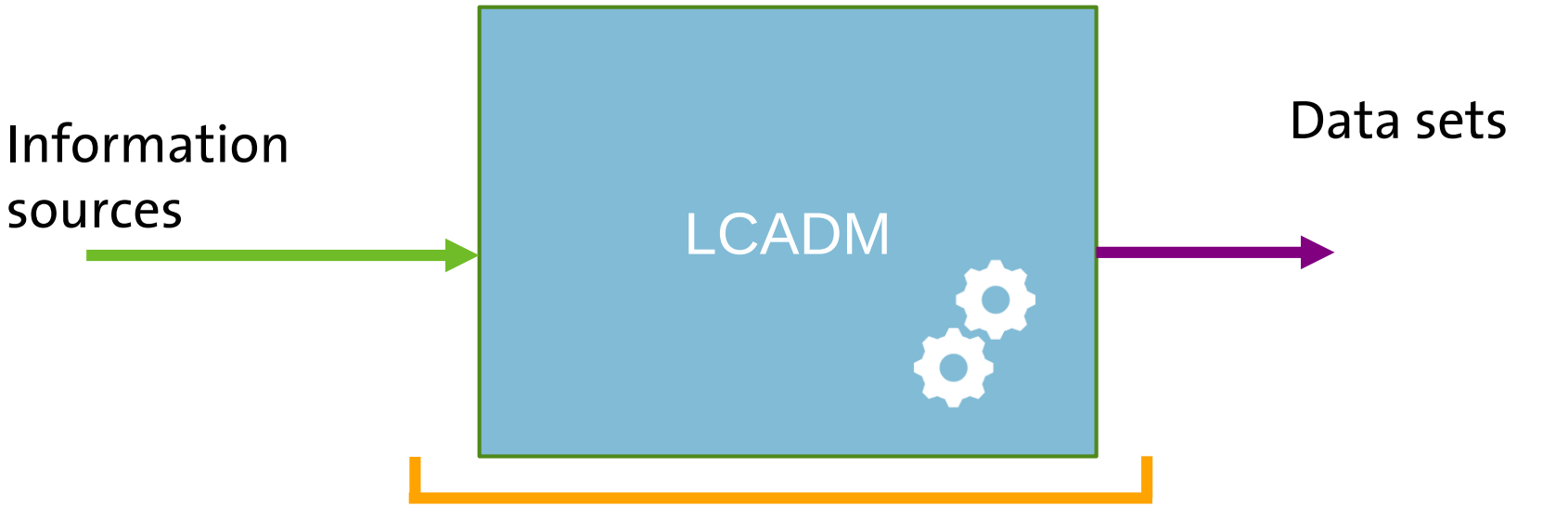
Flow connections and relations (

C-content in fuel is linked to CO_2 emissions
Mass balance for processes
Energy balance for processes

Geography patterns for agriculture

..

The LCA data machine, a simple picture



The LCA data machine, sources

- entsoe energy statistics
- wikipedia
- emission statistics
- (literally, hundreds)...

No use of LCA data sources → instead, raw data outside of the LCA domain.

The LCA data machine, workflow

- Starting point: request for data set, with specification to understand fitness for purpose
- Seed: Process type pattern,
- Completed and refined with constraints and rules, for information from sources
- Fitness for purpose and Shannon entropy calculated
- ..until sufficient

Result is **a core dataset**.

Expert interaction for refinement and review of rules and constraints

The LCA data machine, workflow

- Starting point: request for data set, with specification to understand fitness for purpose
- Seed: Process type pattern,
- Completed and refined with constraints and rules, for information from sources
- Fitness for purpose and Shannon entropy calculated

- ..until sufficient

Result is **a core dataset**.

Expert interaction for refinement and review of rules and constraints

The LCA data machine, workflow, 2

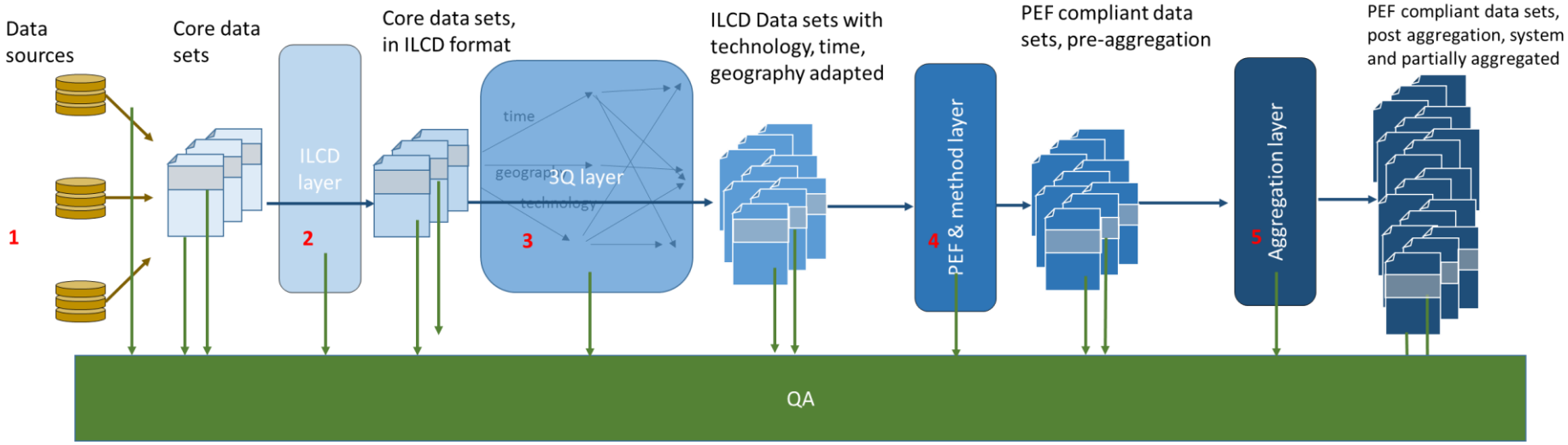
Specific schema requirements are applied on the core data sets after their initial creation

- Use of specific nomenclature
- Specific modelling requirements

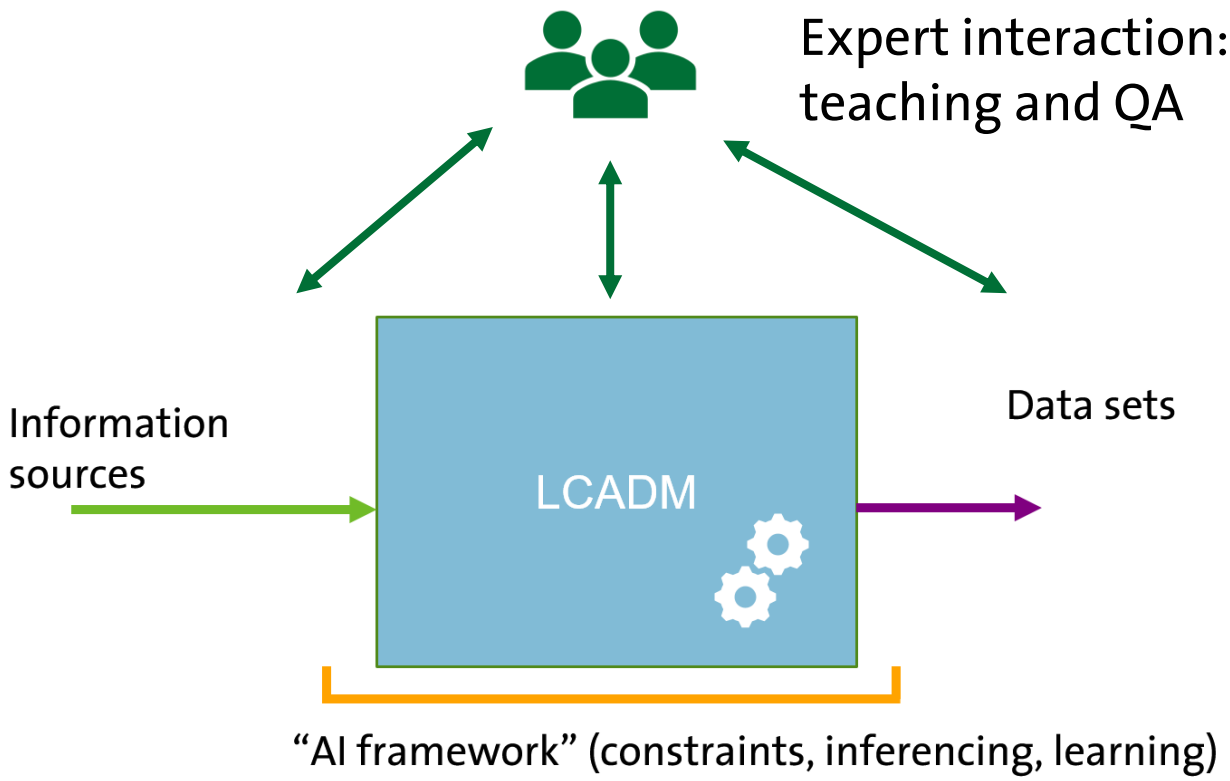
→ This keeps the system flexible; requirements change over time e.g.

- Processing of core data sets, e.g. to migrate to other countries, is also possible

The LCA data machine, one more detailed example



The LCA data machine, expert interaction



3 Some examples

The LCA data machine, some example applications

- Creating a new data set for a specific coal power plant in Poland
- “migrating” a data set to another country (truck transport data set for the US based on a truck transport data set for average Europe)
- Creating a PEF compliant data set from a core data set
- Adaptation of the existing rule and source basis to a specific application (country / product range)

4 Status and outlook, discussion

LCADM: Status

- Some delay due to selection and configuration of the correct AI “portfolio”
- Now really promising
- Rules and constraints for the 12 archetype pattern processes developed
- Machine has the ability to create data sets from scratch and to migrate data sets from one geographic setting to another
- Currently further extended
- A first public version expected end of 2017

LCADM: Invitation & outreach

- **We invite interested experts** to contribute rules and sources, and to act as reviewers
  - Paid-for review and work
  - Detailed announcement start of June 2017
  - If you are interested, send us a note or contact me directly
- **We invite LCA data creation initiatives** to get into contact
  - LCADM as “sparrings partner” for traditional data creation
  - LCADM as starting point, or one input source

LCADM: Outlook

- **Plan: establish LCADM as one LCA data source**
- **Not as competition to the existing LCA data sources, but as addition,** to satisfy a broader need, to create data sets faster, for more widespread uses

LCADM: Outlook

- **Plan: establish LCADM as one LCA data source**
- **Not as competition to the existing LCA data sources, but as addition,** to satisfy a broader need, to create data sets faster, for more widespread uses
- Additional benefit of the LCADM: transparent rules for data sets, in difference to ad-hoc, possibly not documented export decisions existing possibly for “traditional” LCA data sets

Final, personal note

- **LCA data set creation today is** often “somehow strange”, approaches only accepted in our niche community:
  - Data sets that are declared as being representative, but they are not
  - Data sets published as being compliant to a schema, but they are not
  - Data sets published which are even technically inconsistent (invalid formulas, missing links, …)
- **We should all work together to overcome this. Using more advanced, smarter technology is one key**

GreenDelta

sustainability consulting + software

Merci!

Contact: Dr. Andreas Ciroth

GreenDelta GmbH

Müllerstrasse 135, 13349 Berlin, Germany

ciroth@greendelta.com

www.greendelta.com