

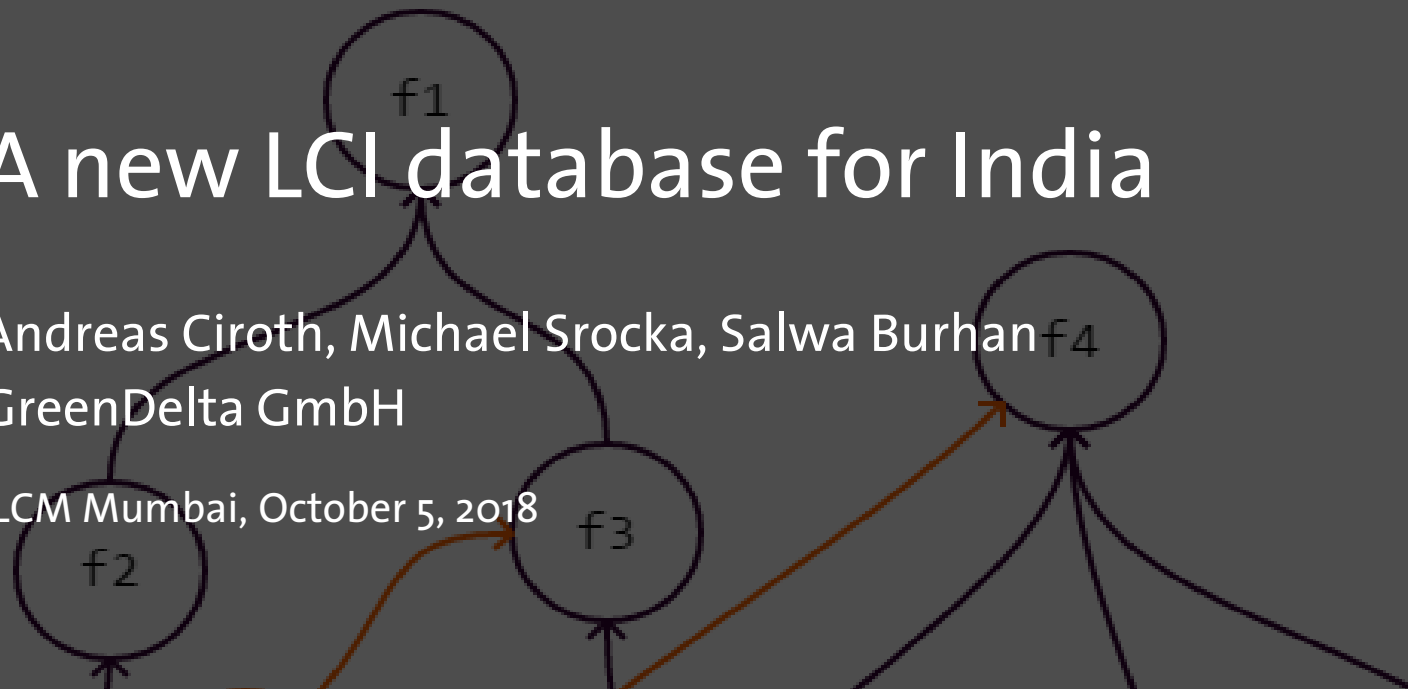
GreenDeLTA

sustainability consulting + software

A new LCI database for India

Andreas Ciroth, Michael Srocka, Salwa Burhan
GreenDelta GmbH

ILCM Mumbai, October 5, 2018



A new LCI database for India

1. Motivation & context
2. Principles of the LCA data machine
3. (towards) an Indian LCI database
4. Status and outlook

1 A motivation, context

A motivation

(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

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- Use of outdated information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)

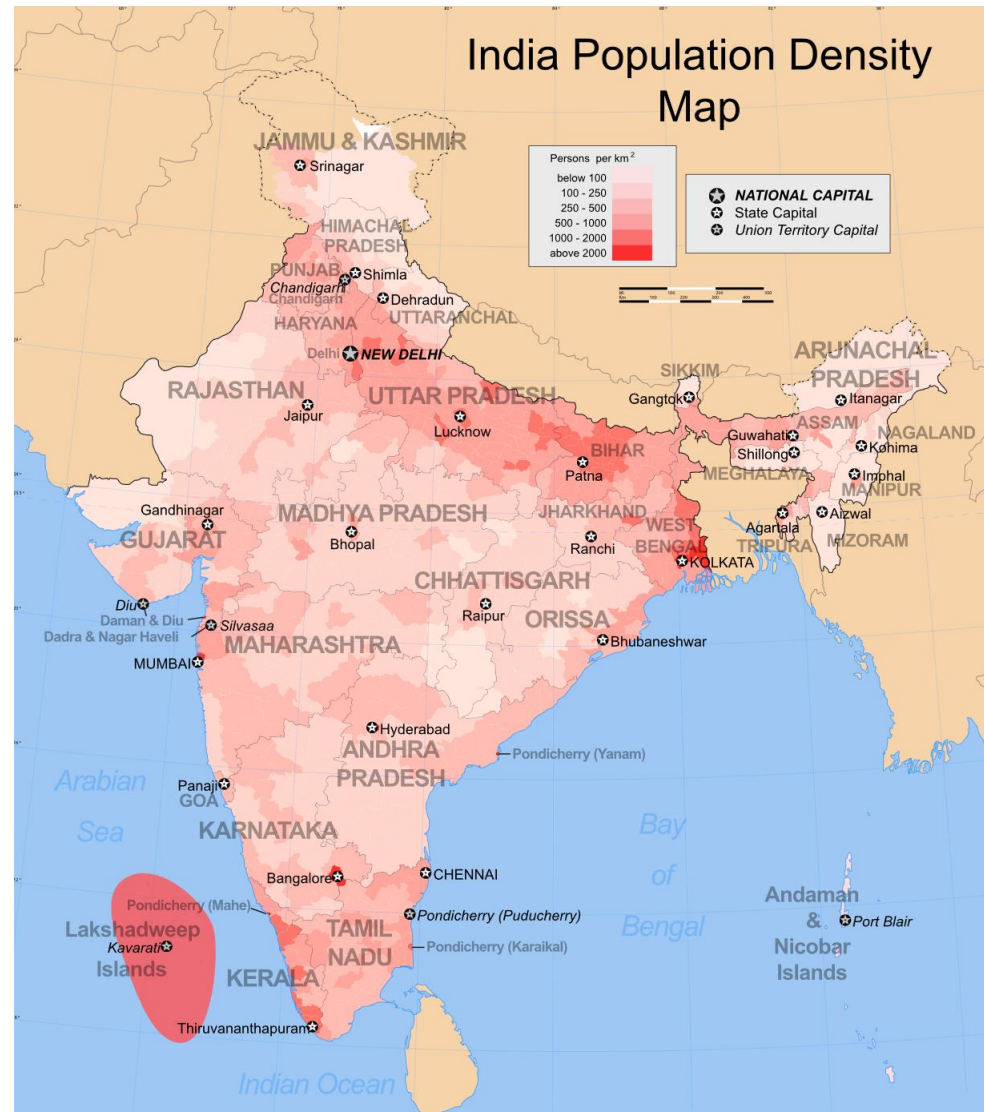
=

- 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“, 2016-10/2018



Bundesministerium
für Wirtschaft
und Energie

India



India & LCA

- Rising interest and recognition for LCA
- Production processes, climate, products different -> datasets from “elsewhere” not fully fitting
- Currently (and since some time) initiatives to develop datasets for India (e.g., ecoinvent, GaBi), ...
- ...using the approaches that were successful 20 years ago in Europe (e.g.), ...
- **which are admittedly slow and require lot of effort**

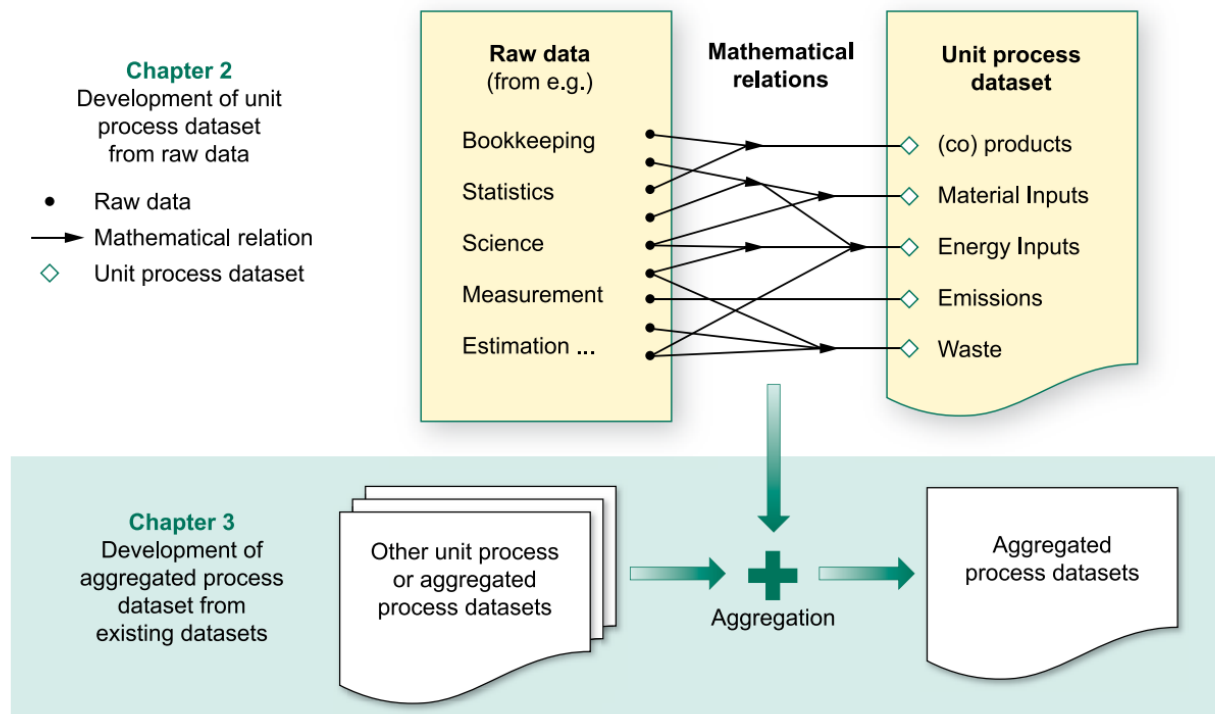
India & LCA

- Rising interest and recognition for LCA
 - Production processes, climate, products different -> datasets from “elsewhere” not fully fitting
 - Currently (and since some time) initiatives to develop datasets for India (e.g., ecoinvent, GaBi), ...
 - ...using the approaches that were successful 20 years ago in Europe (e.g.), ...
 - **which are admittedly slow and require lot of effort**
- is there a chance for new approaches that are faster and cheaper (and maybe even better)

2 Principles of the LCA data machine

The LCA data machine, principles

- 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

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

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

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

The LCA data machine, principles

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

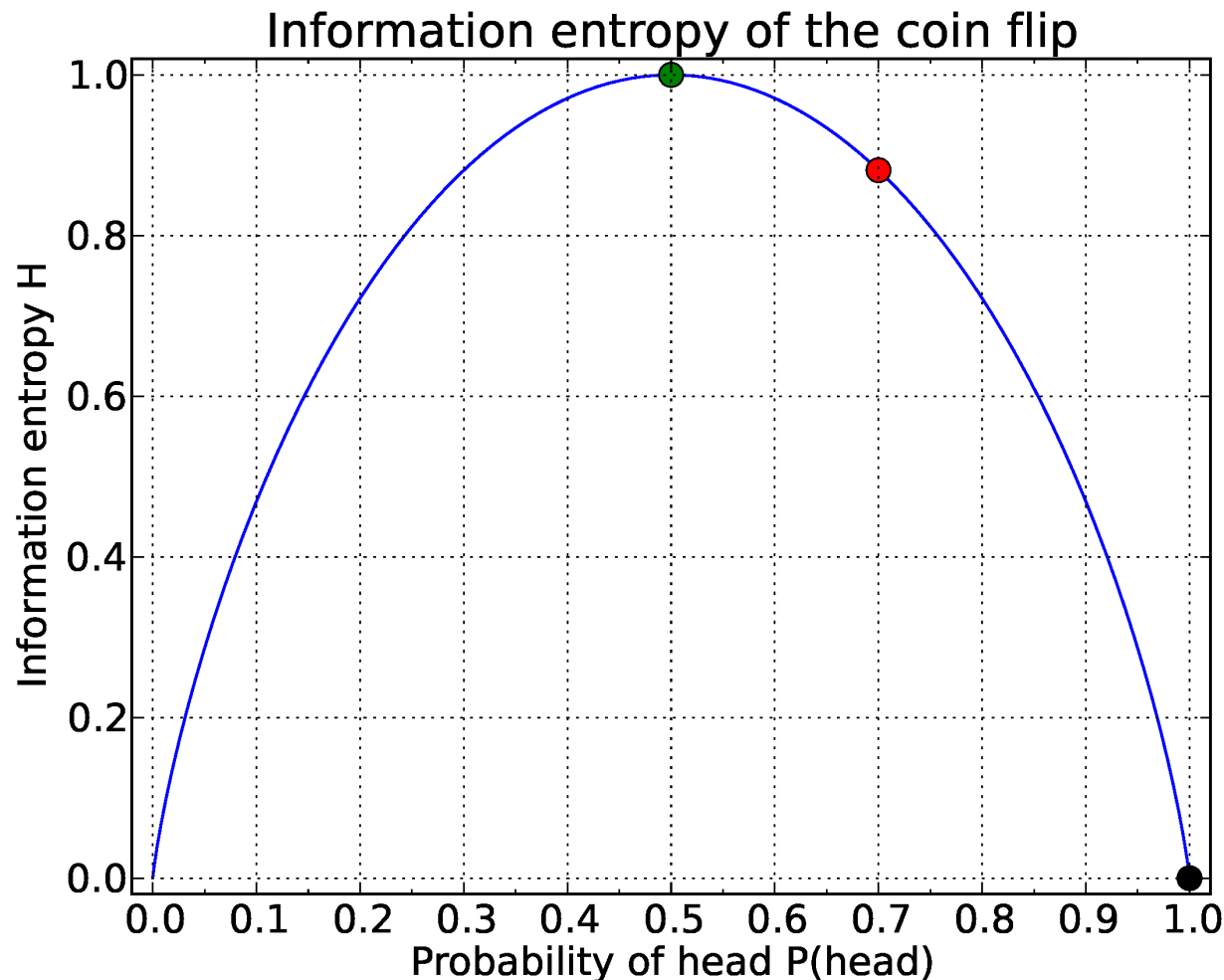
$$H(A) = - \sum_{i=1}^n p_i * \log_2 p_i$$

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

- 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

Use of the UN GLAD metadata descriptors to describe fitness for purpose

- ~ 30 different descriptors

API field	field format	field content	default	comment
refId:	String	The unique identifier of the data set *required		
name:	String	The name of the data set *required		
dataSetUrl:	String	A url to download the complete data set *required		
category:	String	The category of the data set. The value will be automatically build from the elements in the 'categories' field, concatenated with a slash (/). e.g. categories = ['Emission to air', 'Unspecified'] => category = 'Emission to air/Unspecified'		
description:	String	The description of the data set		
technology:	String	A description of the technology used in the data set		
format:	String	Enum: ECOSPOLD1, ECOSPOLD2, ILCD, JSON-LD, OTHER, UNKNOWN		
location:	String	The location of the data set		
dataProvider:	String	The name of the provider of the data set		
supportedNomenclatures:	String	The nomenclatures, the data set is compliant to		
lciaMethods:	[string]	A list of supported LCIA methods		
categoryPaths:	[string]	(Internally) used to build a tree like category structure. This value will be automatically calculated from the elements in the 'categories' field. e.g. categories = ['Emission to air', 'Unspecified'] => categoryPaths = ['Emission to air', 'Emission to air/Unspecified']		
unspscPaths:	[string]	(Internally) used to build a tree like structure for the unspsc code. This value will be automatically calculated from the 'unspscCode' field. e.g. unspscCode = '50454302' => unspscPaths = ['50', '5045', '504543', '50454302']		
co2pePaths:	[string]	(Internally) used to build a tree like structure for the co2pe code. This value will be automatically calculated from the 'co2peCode' field. e.g. co2peCode = '1.1.1' => co2pePaths = ['1', '1.1', '1.1.1']		
processType:	string	Enum: UNIT, PARTIALLY_AGGREGATED, FULLY_AGGREGATED, BRIDGE, UNKNOWN	Default: UNKNOWN	
representativenessType:	string	Enum: SCIENTIFIC, EXPERT_BASED	Default: EXPERT_BASED	
modelingType:	string	Enum: ATTRIBUTIONAL, CONSEQUENTIAL, UNKNOWN	Default: UNKNOWN	
multifunctionalModeling:	string	Enum: PHYSICAL, ECONOMIC, CAUSAL, SYSTEM_EXPANSION, NONE, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
biogenicCarbonModeling:	string	Enum: OMITTED, DISTINGUISHED, AGGREGATED, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
endOfLifeModeling:	string	Enum: CUT_OFF, PHYSICAL_APOS, ECONOMIC_APOS, SUBSTITUTION, OTHER, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
waterModeling:	string	Enum: AMOUNTS, AMOUNTS_AND_AVAILABILITY, AMOUNTS_AND_QUALITY, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
		Enum: INCLUDED, AND_DISTINGUISHED_INCLUDED, AND_NOT_VISIBLE, NOT_INCLUDED		

Use of the UN GLAD metadata descriptors to describe fitness for purpose

- ~ 30 different descriptors

emissionModeling:	string	Enum: INCLUDED_AND_DISTINGUISHED, INCLUDED_AND_NOT_VISIBLE, NOT_INCLUDED, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	(long term emissions)
carbonStorageModeling:	string	Enum: INCLUDED_AND_DISTINGUISHED_CORRECTION, INCLUDED_AND_DISTINGUISHED_OTHER, INCLUDED_AND_NOT_VISIBLE, NOT_INCLUDED, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
sourceReliability:	string	Enum: MEASURED_VERIFIED, PARTLY_MEASURED_VERIFIED, PARTLY_MEASURED_PARTLY_ESTIMATED, ESTIMATED_QUALIFIED, ESTIMATED_UNQUALIFIED	Default: ESTIMATED_UNQUALIFIED	
aggregationType:	string	Enum: HORIZONTAL, VERTICAL, COMBINED, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
reviewType:	string	Enum: INTERNAL, EXTERNAL, PANEL, UNKNOWN, NONE	Default: NONE	
reviewSystem:	string	Enum: ILCD, PEF, GHG, LCA_UN, OTHER, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	
unspscCode:	string	A UNSPSC process code categorizing the data set		
co2peCode:	string	A CO2PE product code identifying the product of the data set		
copyrightHolder:	string	The owner of the copyright of the data set if applicable		
license:	string	The license the data set is released under		
contact:	string	A contact person for information on the data set		
categories:	[string]	The categories of the data set as array (one entry per child category)		
reviewers:	[string]	A list of the names of the reviewers of the data set		
validFrom:	integer (int64)	The start of the validity of the data set in milliseconds since 01/01/1970 (unix-time tims 1000)		
validUntil:	integer (int64)	The end of the validity of the data set in milliseconds since 01/01/1970 (unix-time tims 1000)		
validFromYear:	integer (int32)	The year of the start of the validity of the data set, will be taken from validFrom if not set		
validUntilYear:	integer (int32)	The year of the end of the validity of the data set, will be taken from validUntil if not set		
latitude:	number (double)	The latitude of the geography of the data set		
longitude:	number (double)	The longitude of the geography of the data set		
completeness:	number (double)	The percentage of flows according to nomenclature	Default: 100	
amountDeviation:	number (double)	The deviation in mass and energy balance		
representativenessValue:	number (double)	The percentage of variation coefficient, $s/(\text{arithm mean})$		
copyrightProtected:	boolean	Indicates if the data set is copyright protected		
free:	boolean	Indicates if the data set is available for free		
publiclyAccessible:	boolean	Indicates if the data set can be downloaded from the given dataSetUrl without further login		

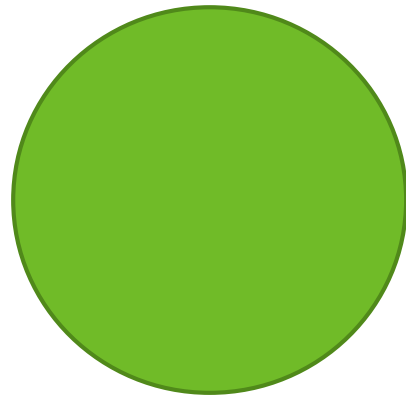
Use of the UN GLAD metadata descriptors to describe fitness for purpose

- ~ 30 different descriptors

validFrom:	integer (int64)
validUntil:	integer (int64)
validFromYear:	integer (int32)
validUntilYear:	integer (int32)

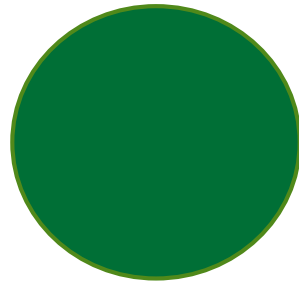
Shannon entropy vs fitness for purpose

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



Data set older than 2016

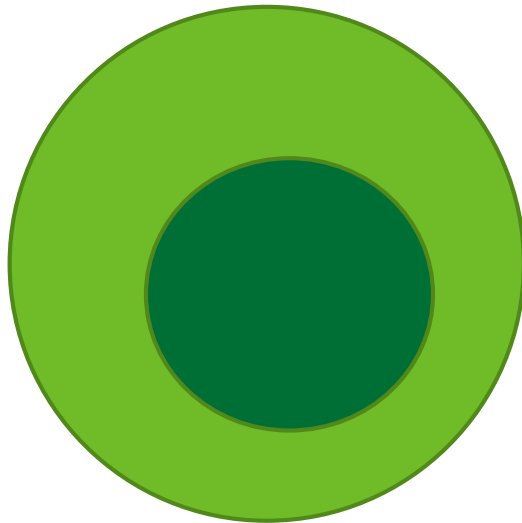
- More specific information



Data set from 2005

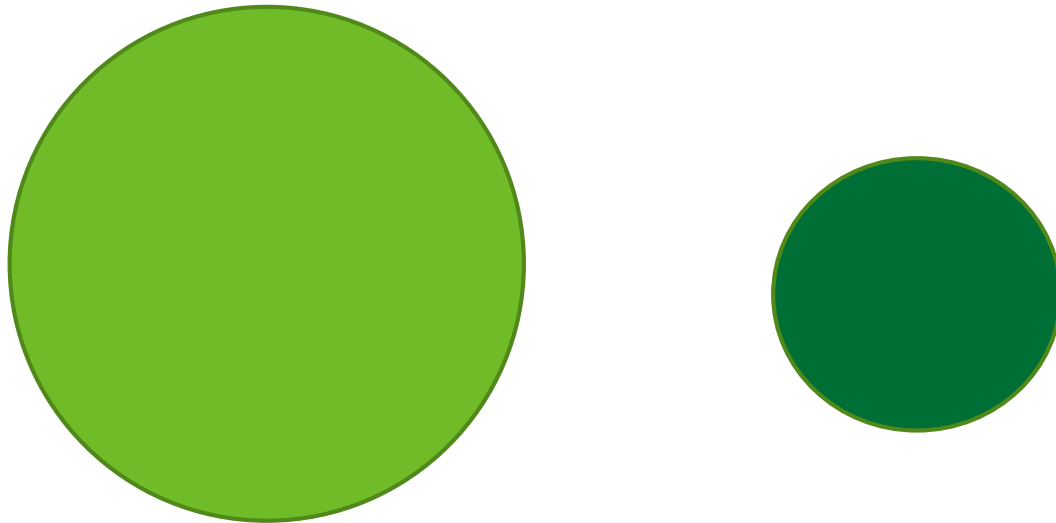
Shannon entropy vs fitness for purpose

- **Case 1:** broader information in scope, more detailed information also in scope
Shannon entropy \searrow , fitness for purpose \nearrow , improvement



Shannon entropy vs fitness for purpose

- **Case 2:** broader information in scope, more detailed information not in scope
Shannon entropy \nearrow , fitness for purpose \downarrow , degradation



The LCA data machine, 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 CO₂ emissions
Mass balance for processes
Energy balance for processes

Geography patterns for agriculture

..

The LCA data machine, a bit more informatics..

- Machine written in Clojure and meant to be used in an interactive way, using REPL (read-eval-print-loop)
- an own, internal, JSON format developed to store the information

Ur 0.1.0-SNAPSHOT

Project	Public Vars
Index	-main
Namespaces	broader?
ur	direct-broader
core	direct-narrower
exchanges	exchanges
graphviz	flow
infer	make-graph
io	narrower?
isets	read-graph
spold	search
	traverse
	traverse-broader

read-graph
(read-graph folder)
Read the data from the given folder creates the graph using the make-graph function.

search
(search graph flow-name)
Search the graph for flows that have the given name in their labels or synonyms, e.g.:

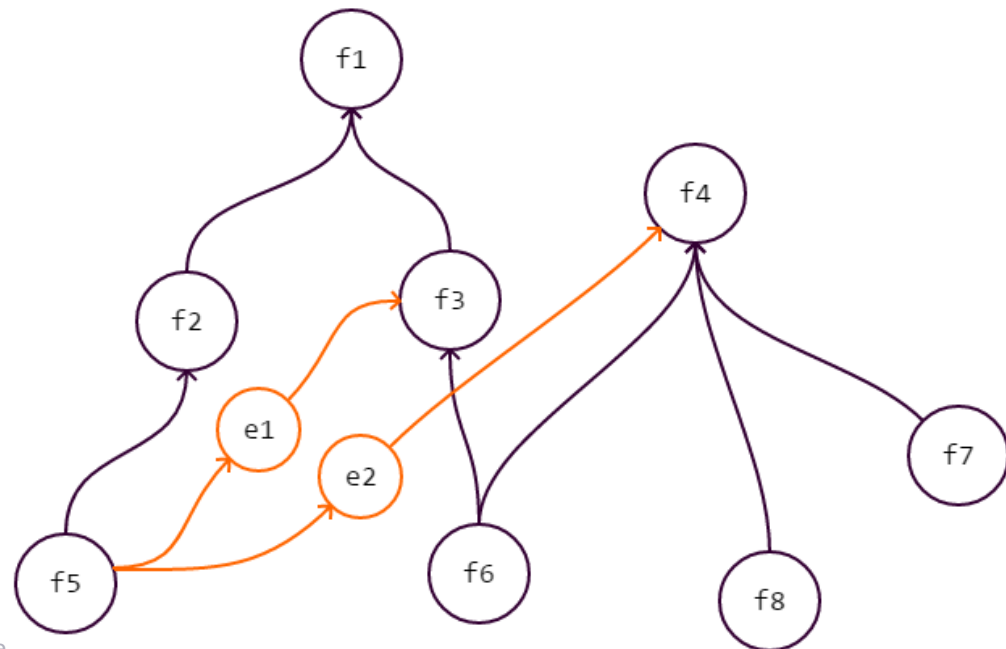
```
(search g "milk")
```

traverse
(traverse graph flow-id f)

traverse-broader
(traverse-broader graph flow-id f)

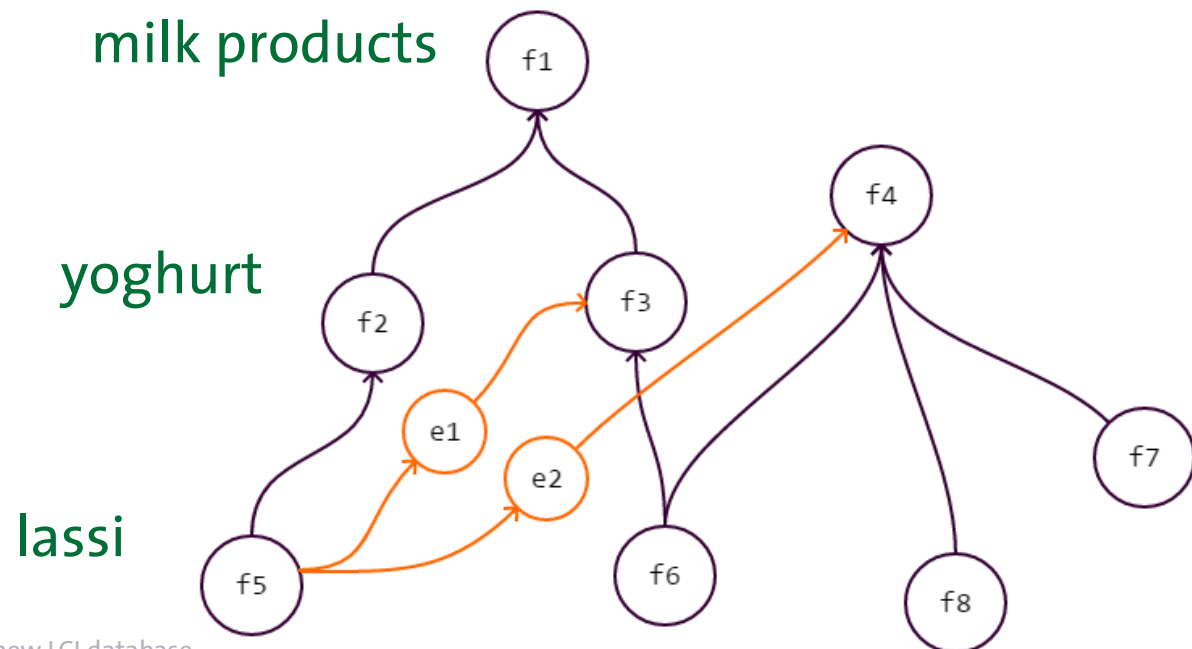
The LCA data machine, usage

- 1) Start, and refine (12 process archetypes)
- 2) Inference graphs: flows and exchanges (i.e, connections between flows), building on the information from various sources, hierarchically organised, in multi dimensions

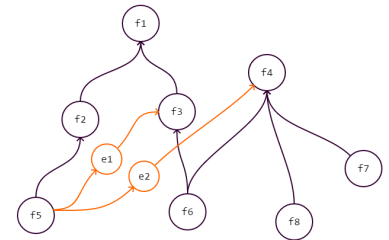


The LCA data machine, usage

- 1) Start, and refine (12 process archetypes)
- 2) Inference graphs: flows and exchanges (i.e, connections between flows), building on the information from various sources, hierarchically organised, in multi dimensions



The LCA data machine, usage



- 1) Start, and refine (12 process archetypes)
- 2) Inference graphs: flows and exchanges (i.e, connections between flows), building on the information from various sources
- 3) Start “somewhere” on the graph, and collect information along the graph (go up, higher hierarchies, go down again, also in other branches, traverse)
- 4) Replace existing information for the process with better
- 5) Very powerful and flexible and lightweight programming language which can “consume” a broad variety of data

The LCA data machine, status

- 1) “infrastructure” is set up
- 2) end to end pipeline works (from raw data in acyclic graphs to inferred datasets in EcoSpold1 XML and JSON-LD → import into openLCA)
- 3) Data sources and rules refined and extended

Welcome | P Gimped yarn - GLO

Inputs/Outputs: Gimped yarn

Inputs

Flow	Category	Amount Unit	Costs/Reve...	Uncertainty	Avo
aluminium melting furnace - RER - RER		2.0E-9 Item(s)		none	
aluminium oxide, at plant - RER - RER		0.0214 kg		none	
aluminium oxide, plant - RER - RER		2.5E-11 Item(s)		none	
aluminium scrap, new, at plant - RER - RER		1.01 kg		none	
aluminium scrap, old, at plant - RER - RER		1.03 kg		none	
aluminium sulphate, powder, at plant - RER - RER		1.0907441775 kg		none	
aluminium, production mix, at plant - RER - RER		390000.0 kg		none	
aluminium, production mix, cast alloy, at plant - RER - RER		0.965 kg		none	
aluminium, production mix, wrought alloy, at plant - RER - RER		5.7409E-4 kg		none	
ammonia, liquid, at regional storehouse - CH - CH		2.9936787499999999 kg		none	
ammonia, liquid, at regional storehouse - RER - RER		0.04082985 kg		none	
ammonium bicarbonate, at plant - RER - RER		0.0013088 kg		none	
ammonium carbonate, at plant - RER - RER		0.00121 kg		none	

Outputs

Flow	Category	Amount Unit	Costs/Reve...	Uncertz
Elementary flows / Emission to air / low population density / Beryllium		1.14094696777... kg		none
Elementary flows / Emission to air / low population density / Boron		1.69728643689... kg		none
Elementary flows / Emission to air / low population density / Bromine		1.861825E-5 kg		none
Elementary flows / Emission to air / low population density / Butadiene		9.62724999999... kg		none
Elementary flows / Emission to air / low population density / Butane		9.13170000000... kg		none
Elementary flows / Emission to air / low population density / Cadmium		9.1602E-9 kg		none
Elementary flows / Emission to air / low population density / Calcium		2.4060875E-6 kg		none
Elementary flows / Emission to air / low population density / Carbon dioxide, biogenic		0.191205 kg		none
Elementary flows / Emission to air / low population density / Carbon dioxide, fossil		0.2844525 kg		none
Elementary flows / Emission to air / low population density / Carbon dioxide, land transformation		0.00248965 kg		none

3 Application for India

An LCI database for India, data sources

- Wikipedia (subset)
- Public data (data.gov.in)
- Meteorological data
- LCA Data, e.g. case studies (!)

An LCI database for India, approach

- Refining and adjusting data, for India (e.g.:
 - climate (even more local, not entire India)
 - different BOM / inputs (pesticides in agriculture, ...)
 - different efficiencies, different amounts
- Building new datasets

4 Status & outlook

Status

- **LCA data machine, as “engine” for the database:**
 - infrastructure set up
 - end to end processing works
 - results are promising
- **Indian database:** Very interesting application case
- Promises to be fast and able to deliver many datasets in good quality, fit for purpose (!)

Invitation & outreach

- **We invite interested experts** to contribute rules and sources, and to act as reviewers
 - Recognised review and work
 - If you are interested, send us a note or contact me directly
- **We invite LCA data creation initiatives** to get into contact
 - Application as “sparrings partner” for traditional data creation
 - As starting point, or one input source
 - Also feedback on how to provide data

Outlook

- **Plan: establish the LCA data machine and Indian database 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**

Outlook

- **Plan: establish the LCA data machine and Indian database 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 LCA data machine: transparent rules for data sets, in difference to ad-hoc, possibly not documented export decisions existing possibly for “traditional” LCA data sets

GreenDelta

sustainability consulting + software

Thank you!

Contact: Dr. Andreas Ciroth
GreenDelta GmbH
Müllerstrasse 135, 13349 Berlin, Germany
ciroth@greendelta.com
www.greendelta.com

