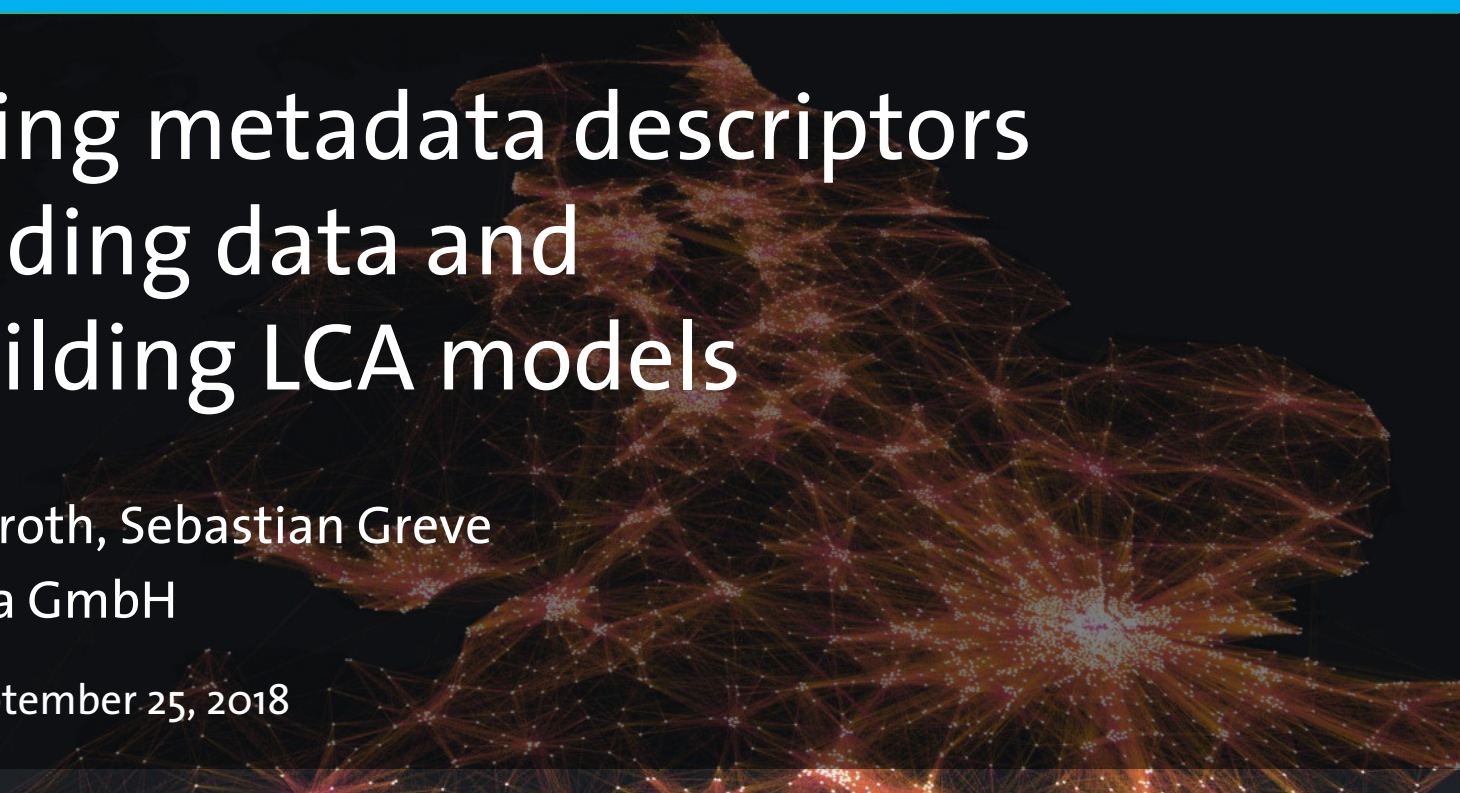# greenDelTa

sustainability consulting + software

Applying metadata descriptors for finding data and for building LCA models

Andreas Ciroth, Sebastian Greve

GreenDelta GmbH

LCA XVIII, September 25, 2018

Applying metadata descriptors for finding data and for building LCA models
Fitness for purpose based autolinking in heterogenous data pools

Applying metadata descriptors for finding data and for building LCA models
Fitness for purpose based autolinking in heterogenous data pools

1. UN GLAD descriptors for assessing fitness for purpose

2. Application 1: Searching for best fitting datasets

3. Application 2: Linking in LCA product systems

   1. Concept

   2. Application case

   3. Next steps

4. Discussion: what would this concept change for LCA?

# 1 UN GLAD Descriptors for assessment of data quality

UN GLAD descriptors for assessing fitness for purpose (= data quality)

- Result of an international Working Group 2015 – 2017

- About 30 indicators that describe aspects of data quality for LCA inventory data sets

- Idea: capture content and LCA modelling details of LCA datasets

- Data quality is called fitness for purpose

- Recently implemented in an API and search engine, open source
  https://github.com/GreenDelta/glad-rest-service

- API integrated in a central website,
  globallcadataaccess.org

UN GLAD descriptors for assessing fitness for purpose (= data quality)

- Common structure for each descriptor:

  - goal (what the dataset ideally represents)

  - representation / value (what the dataset actually represents)

  - conformance (delta between goal and representation)

  - representation and conformance can be reviewed, the goal cannot be reviewed (since it is given by the application / user)

UN GLAD descriptors for assessing fitness for purpose (= data quality)
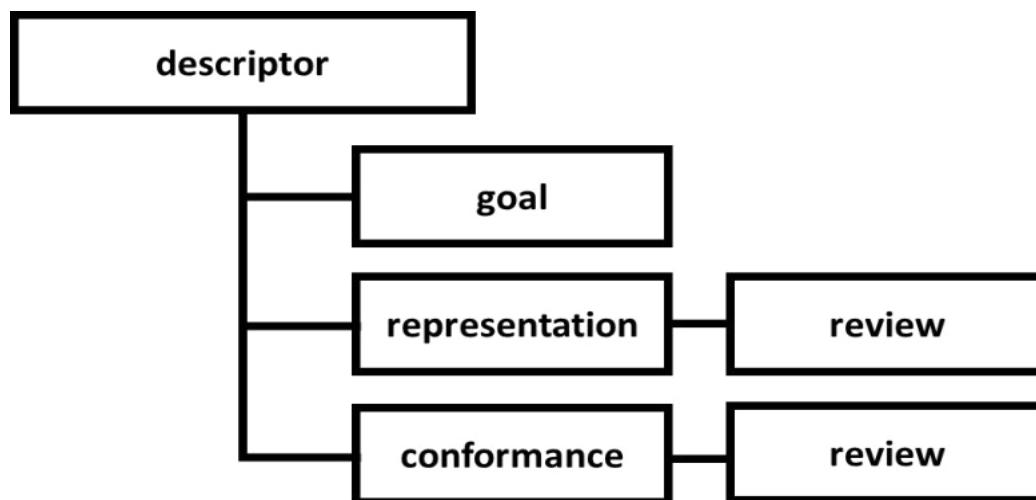
Ciroth, A., Arbuckle, P., Cherubini, E., Ugaya, C., Edelen, A.: Core meta-data descriptors and guidance on populating descriptors WG3 of the Global Life Cycle Data Access Network (GLAD), Release version 1.0 (internal version 11.3), June 2017
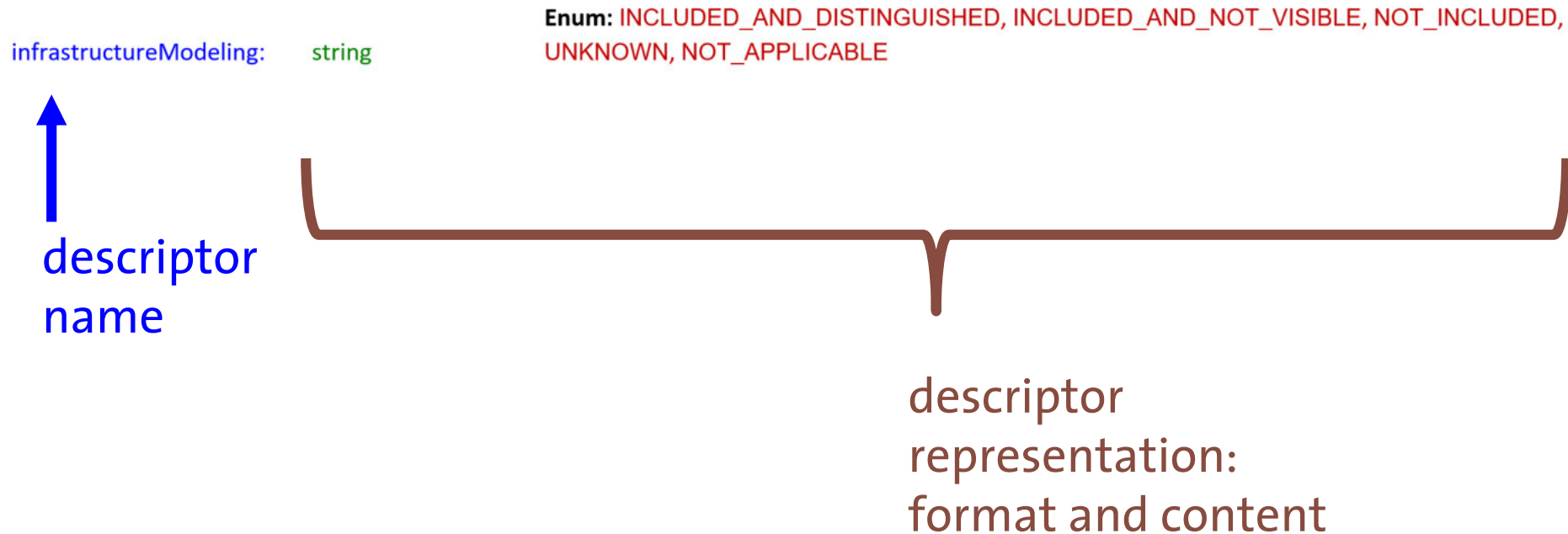
UN GLAD descriptors for assessing fitness for purpose (= data quality)

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, concatinated 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	
infrastructureModeling:	string	Enum: INCLUDED_AND_DISTINGUISHED, INCLUDED_AND_NOT_VISIBLE, NOT_INCLUDED, UNKNOWN, NOT_APPLICABLE	Default: NOT_APPLICABLE	

UN GLAD descriptors for assessing fitness for purpose (= data quality)

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 infomation 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/(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		

UN GLAD descriptors for assessing fitness for purpose (= data quality)

infrastructureModeling: string

Enum: INCLUDED_AND_DISTINGUISHED, INCLUDED_AND_NOT_VISIBLE, NOT_INCLUDED, UNKNOWN, NOT_APPLICABLE

descriptor name

descriptor representation: format and content

2 Application case 1: finding best fitting datasets

Finding datasets best fitting for purpose (i.e., of best quality), application case 1

- Searching sugar-related datasets in LCA databases

Search box: sugar | Search | Options

Database more
- Agri-footprint 282
- ecoinvent 233
- soca 227
- Social Hotspots 226
- EuGeos' 15804-IA 223
- exiobase 96
- GaBi 92
- PSILCA 78
- ESU World Food 33
- ProBas 14
- more...

Country more
- Netherlands 165
- Switzerland 161
- Germany 149
- Brazil 146
- France 138
- United States 134

1311 data sets in 1553 ms

Sugar cane, sugar beet (Albania)
Databases: Social Hotspots Database
Category: Social Hotspots DB/Albania
Version (internal): 00.00.000   Location: Albania

Sugar cane, sugar beet (Armenia)
Databases: Social Hotspots Database
Category: Social Hotspots DB/Armenia
Version (internal): 00.00.000   Location: Armenia

Sugar cane, sugar beet (Angola)
Databases: Social Hotspots Database
Category: Social Hotspots DB/Angola
Version (internal): 00.00.000   Location: Angola

GreenDELTA

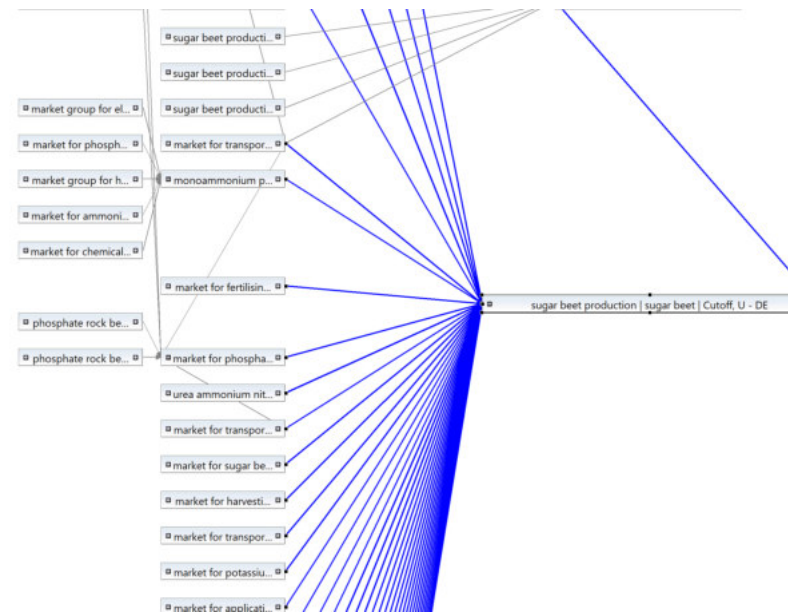# Finding datasets best fitting for purpose (i.e., of best quality), application case 1

- User specifies what he or she needs ($\rightarrow$ "goal")

- Search engine has access to datasets and to descriptors of datasets (to the "value and representation" part)

- Search engine applies conformance assessment for each descriptor

- All conformance results are merged to one fitness for purpose index

- Datasets are presented to users as search result list

- User can then select and download the dataset in a desired exchange format

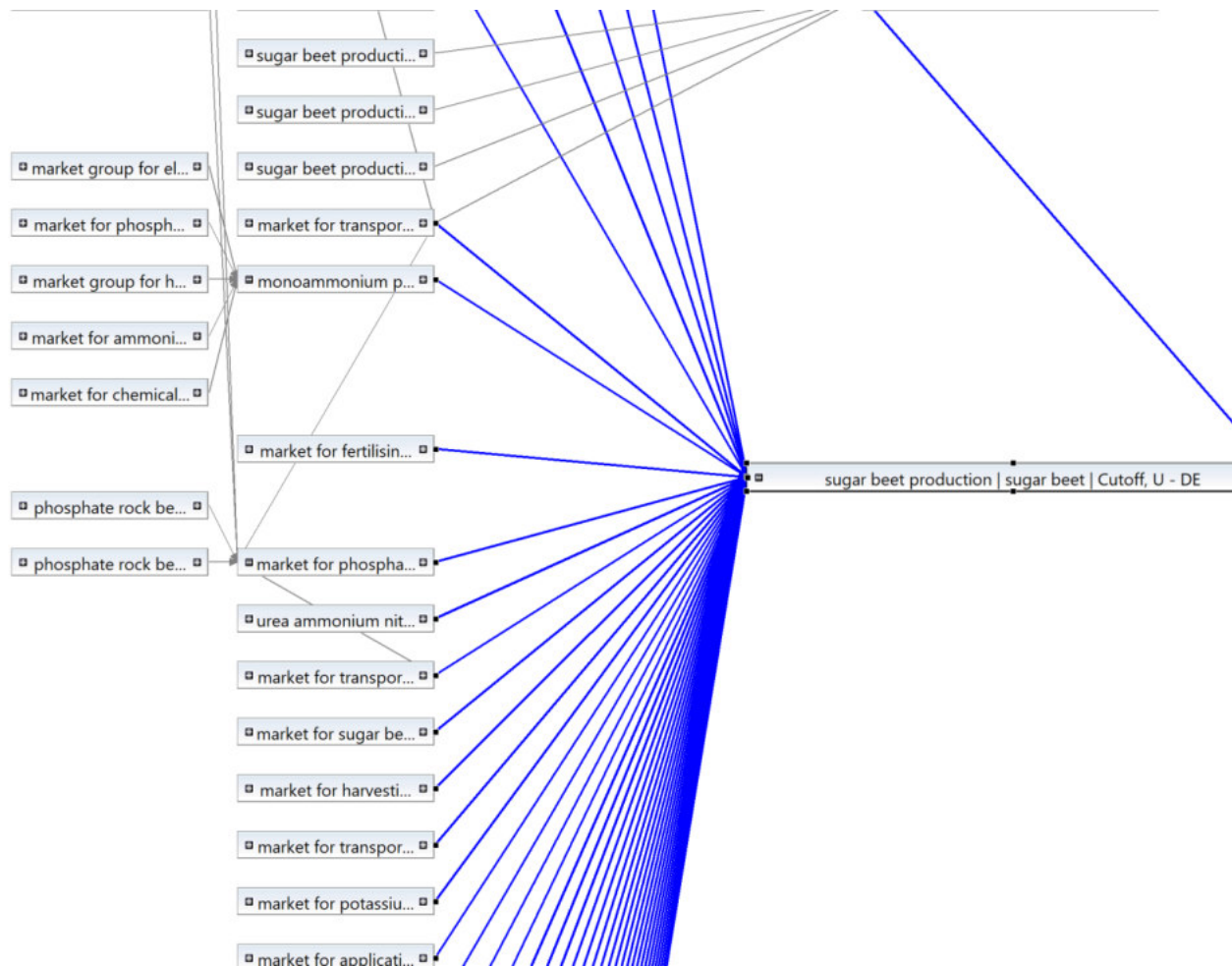# 3 Application case 2: linking in LCA product systems

LCA Product systems

- LCA Product systems are linked life cycles

Product systems: "*collection of unit processes with elementary and product flows, performing one or more defined functions, and which models the life cycle of a product*" [ISO 14040, 3.28]
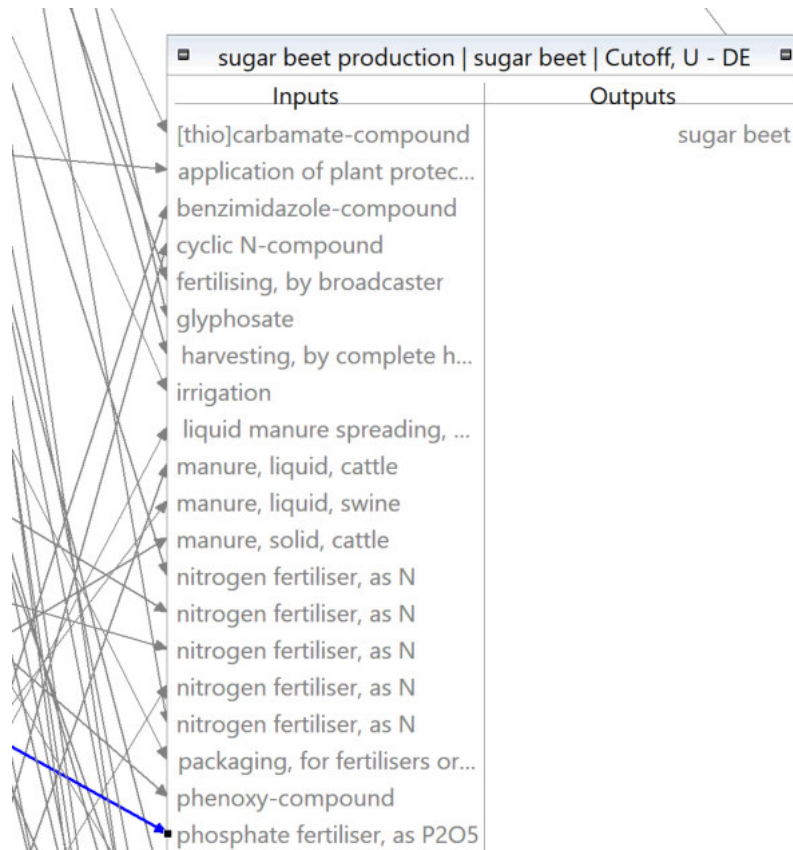
LCA Product systems

sugar beet producti...
sugar beet producti...
market group for el... | sugar beet producti...
market for phosph... | market for transpor...
market group for h... | monoammonium p...
market for ammoni...
market for chemical...
market for fertilisin...
phosphate rock be...
phosphate rock be... | market for phospha...
urea ammonium nit...
market for transpor...
market for sugar be...
market for harvesti...
market for transpor...
market for potassiu...
market for applicati...

sugar beet production | sugar beet | Cutoff, U - DE

LCA Product systems

One example connection: sugar beet production in Germany needs phosphate fertiliser.

GreenDeLTa

LCA Product systems

In order to establish this connection, a process needs to be found that ideally:

- fits to goal and scope of the study (and product system)

  - same reference time

  - same allocation model

  - same way to model biogenic carbon

  - …

- delivers the specific product needed by the sugarbeet process

LCA Product systems

In order to establish this connection, a process needs to be found that ideally:

- fits to goal and scope of the study (and product system)

- delivers the specific product needed by the sugarbeet process

→ i.e., a process needs to be found that best fits for purpose (!)

→ we will call this approach *fitness for purpose autolinking*, FFPA

LCA Product systems

Linking "the next" process in a product system is a bit like taking this or the other road when driving: When you take one direction, this influences the next "available" streets.

LCA Product systems, sugarbeet example

sugarbeet production

input

1 manure

2 N-fertiliser

3 P-fertiliser

...

product system   biogenic c: distinguish
                      reference time: 2015
                      region: Germany
                      ..

2: N-fertiliser	**biogenic c**	**reference time**	**region**	**..**	**technology (process & product)**
reference (from sugarbeet prod.)	distinguish	2015	Germany	..	diammonium phosphate

LCA Product systems, sugarbeet example

	2: N-fertiliser	biogenic c	reference time	region	..	technology (process & product)
reference	reference (from sugarbeet prod.)	distinguish	2015	Germany	..	diammonium phosphate
option a	Calcium ammonium nitrate (CAN, solution), production mix, at plant, technology mix, nitrogen content 26,5% (United States)	distinguish	2013	US		calcium ammonium nitrate, production mix
option b	Urea ammonium nitrate (UAN), production mix, at plant, technology mix, nitrogen content 30% (Germany)	distinguish	2013	Germany		urea ammonium nitrate, production mix
option c	nitrogen fertiliser, use of (Switzerland)	aggregated	2011	Switzerland		nitrogen fertiliser, unspecified [, production mix]

Linking in LCA product systems, next steps

- Provide descriptors for important LCA data sources (also for GLAD, for the search engine)

- Implement the search in LCA software (e.g., openLCA):

  - linking algorithm

  - results display, "data quality" of the links in a product system (instead of the quality of the process)

- Refine concept (i.e., there is a propagation of fitness for purpose in linking: a bad fit in step i cannot entirely be compensated by a good fit in step i+1)

3 Discussion: what would this autolinking change

Discussion points: what would FFPA change?

- status now: existing databases, some (ecoinvent e.g.) with foreseen linking, struggling to capture more data

- combination of different databases hard to decide, typically specific user decisions

- FFPA:

  - agnostic to specific databases, can use any data source available

  - allows to "throw in" new data sources in a data pool, data is considered once it improves the linking, and in the end quality of the model

  → promise: more complete, more realistic model, fitting (much!) better to what the user needs

Discussion points: what would FFPA change?

- FFPA, continued:

  - best suited for unit processes but works with system processes too (fewer linking decisions)

  - works also with already well-linked databases

  - a product system linking is then like a transportation network:

    - motorways (well-linked databases)

    - railways (system processes, a route change is not easily possible)

    - and smaller streets (everything else)

Discussion points: what would FFPA change?

Uber's Kepler application: commute patterns in the UK, http://kepler.gl

FFPA status & call for participation

- Descriptors now in 1st version, more are proposed by the metadata WG; they need to be tested

- E.g. in the Ocelot project, other aspects have been considered but only for ecoinvent (production volume e.g.)

- (1) method, (2) databases, and (3) tool support are needed to make FFPA really powerful

- We plan to extend the idea and extend openLCA and the Nexus databases and our LCA data machine to include the descriptors and to support the FFPA autolinking.

- Any collaboration is welcome!

greenDELTa

sustainability consulting + software

Thank you!

Contact: Dr. Andreas Ciroth

GreenDelta GmbH
Müllerstrasse 135, 13349 Berlin, Germany
ciroth@greendelta.com
www.greendelta.com