GreenDelta

software / data / know-how

Huellas ambientales:
LCA Database creation

Relevamiento de datos y creación de bases de datos (LCI) para análisis de ciclo de vida

Dr Andreas Ciroth

GreenDelta GmbH

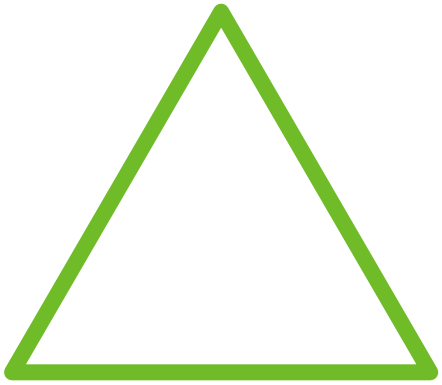
Buenos Aires, 6 November 2018

GreenDelta

SME, 12 people, international (DE, IT, ID, CN), office in Berlin

Sustainability Consultancy and -research

Life Cycle Assessment, resource depletion, Life Cycle Costing, soziale impacts along the life cycle, EPDs



Database development, management and distribution

Software development, open source and other

GreenDelta

- Diverse employees
  (engineers, architects, IT developers, ..)
- International projects and clients:
  - Research projects (EU, German)
  - Institutional projects (German EPA, US EPA, US DA, PTJ, ...)
  - Industrial projects
  - Projects with and for other consultants and also universities, worldwide

GreenDelta, some current projects

- openLCA, world's leading open source LCA software, www.openLCA.org
- UNEP GLAD, data conversion and powerful search for retrieving and providing LCA datasets worldwide
- NEXUS, world's largest repository for LCA and Sustainability data
- PEF remodeling, pilots paper and PV
- PSILCA, world's largest, and transparent, database for Social Life Cycle Assessment
- The LCA data machine

Points for the talk

- Creation of LCA data sets
- Creation of an LCA database

Points for the talk

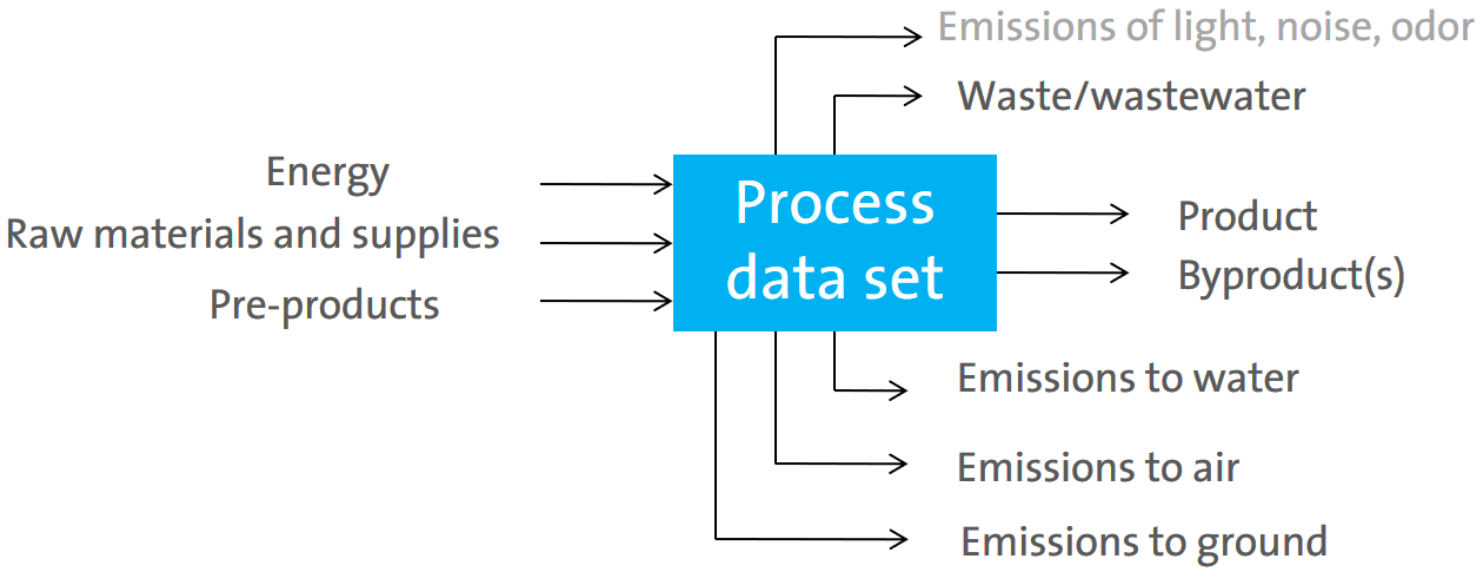
- Creation of LCA data sets
- Creation of an LCA database

... because

- for any LCA study, you will have own datasets and datasets from a database, or from several databases
- and for creating a databases, also datasets need to be created

Creation of LCA datasets

What is an LCA (=Life Cycle Assessment) dataset



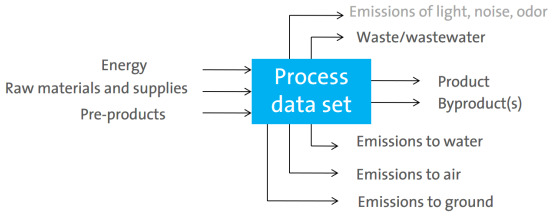
What is an LCA dataset

- an LCA dataset is the smallest modeling unit in a Life Cycle model; the smallest modeling unit is also called “unit process”
- every life cycle model comprises many processes
- the LCA dataset consists basically of
  - input flows (resources, products), and
  - output flows (product, waste, emissions)

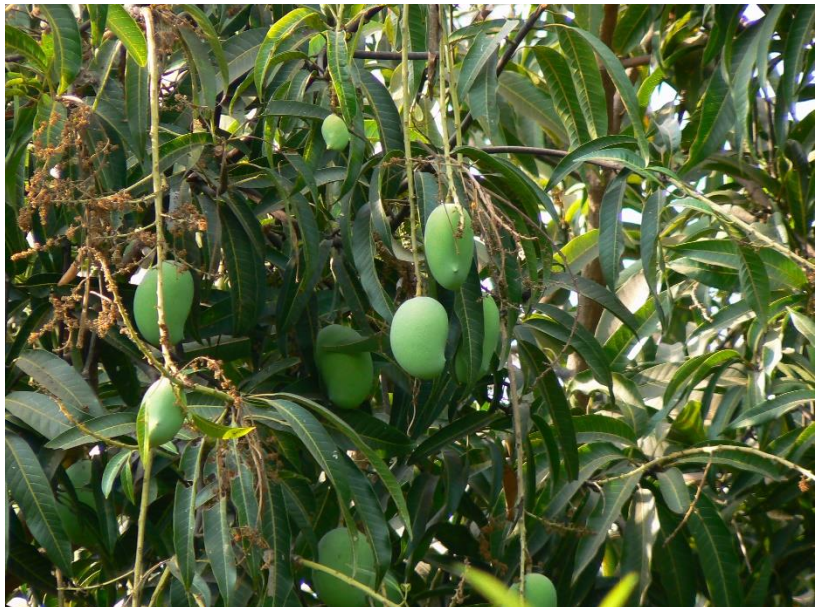
  of one “step” in a life cycle

Primary raw data acquisition

Background

- 

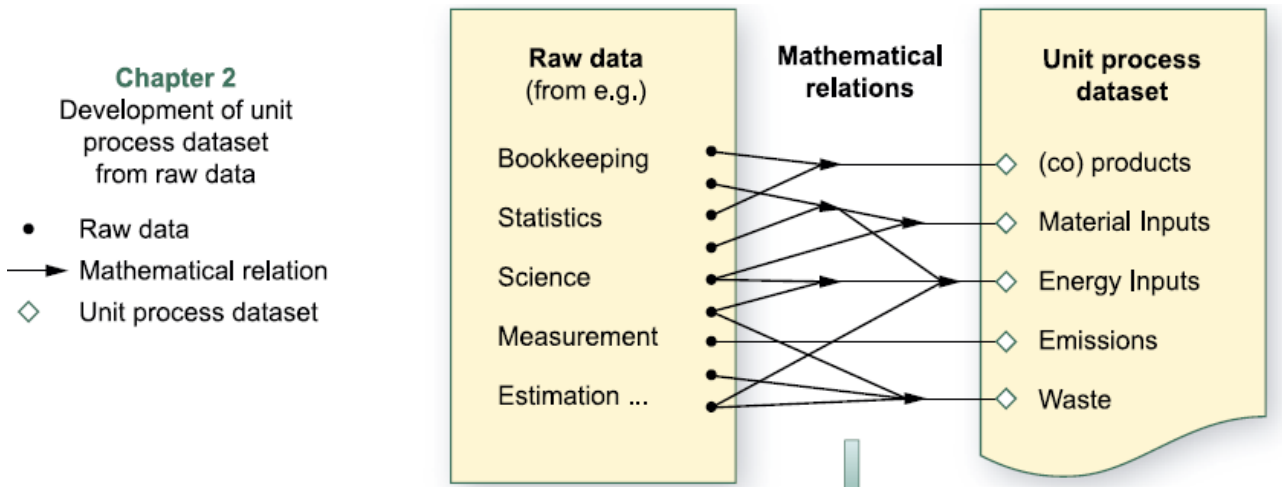
- e.g., Mango production in Argentina



GreenDeLTa

Primary raw data acquisition
Background

- Using “raw data” to create unit process data sets



Wang, H., Ciroth, A., et al.: Development of unit process data sets, chapter 2 in UNEP Shonan Guidance Principles, UNEP 2012

Primary raw data acquisition
An excel sheet for data collection

Process	
Process identification [1]	
Process operator [2]	
Location [3]	
Quantitative reference and unit [4]	
Contact person [5] Address Telephone e-mail	
Process flowsheet [8]	

Inputs	Amount	Unit [9]
Energy source incl. efficency [14]		
Material Inputs [15]		
Service Inputs [16]		

Outputs	Amount	Unit [9]
Product(s) [17]		
Emissions to air [18]		

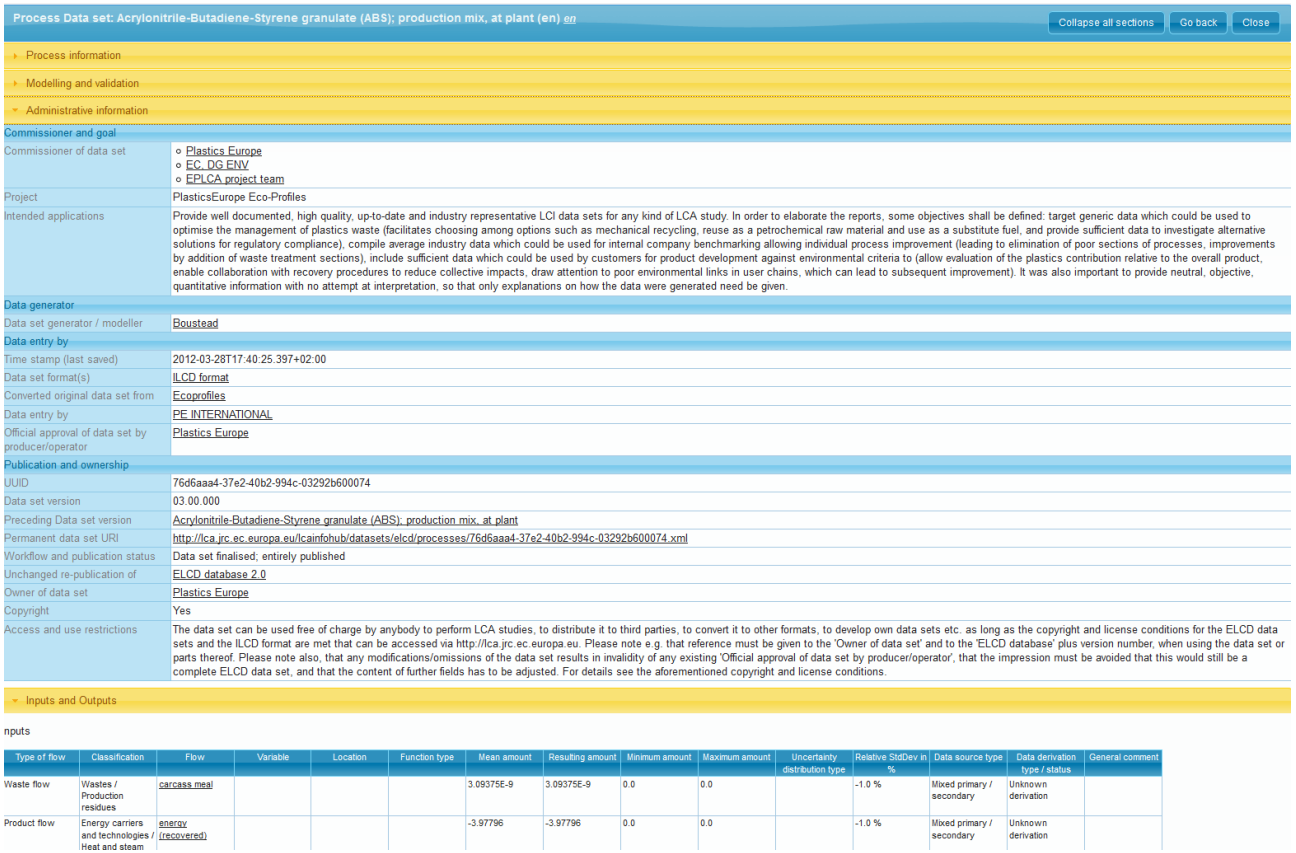
- data collection in LCA is nowadays still often done in a very “traditional” way, using excel sheets for example
- communication & trust very important when collecting information (technical language; language)

GreenDelta

Primary raw data acquisition

Background

- How does a process data set in principle look like (screenshot from a European database)



Process Data set: Acrylonitrile-Butadiene-Styrene granulate (ABS); production mix, at plant (en) *en*

Collapse all sections | Go back | Close

- Process information
- Modelling and validation
- Administrative information

Commissioner and goal

Commissioner of data set	○ Plastics Europe ○ EC, DG ENV ○ EPLCA project team
Project	PlasticsEurope Eco-Profiles
Intended applications	Provide well documented, high quality, up-to-date and industry representative LCI data sets for any kind of LCA study. In order to elaborate the reports, some objectives shall be defined: target generic data which could be used to optimise the management of plastics waste (facilitates choosing among options such as mechanical recycling, reuse as a petrochemical raw material and use as a substitute fuel, and provide sufficient data to investigate alternative solutions for regulatory compliance), compile average industry data which could be used for internal company benchmarking allowing individual process improvement (leading to elimination of poor sections of processes, improvements by addition of waste treatment sections), include sufficient data which could be used by customers for product development against environmental criteria to (allow evaluation of the plastics contribution relative to the overall product, enable collaboration with recovery procedures to reduce collective impacts, draw attention to poor environmental links in user chains, which can lead to subsequent improvement). It was also important to provide neutral, objective, quantitative information with no attempt at interpretation, so that only explanations on how the data were generated need be given.

Data generator

Data set generator / modeller	Boustead

Data entry by

Time stamp (last saved)	2012-03-28T17:40:25.397+02:00
Data set format(s)	ILCD format
Converted original data set from	Ecoprofiles
Data entry by	PE INTERNATIONAL
Official approval of data set by producer/operator	Plastics Europe

Publication and ownership

UUID	76d6aaa4-37e2-40b2-994c-03292b600074
Data set version	03.00.000
Preceding Data set version	Acrylonitrile-Butadiene-Styrene granulate (ABS); production mix, at plant
Permanent data set URI	http://lca.jrc.ec.europa.eu/lcainfohub/datasets/elcd/processes/76d6aaa4-37e2-40b2-994c-03292b600074.xml
Workflow and publication status	Data set finalised; entirely published
Unchanged re-publication of	ELCD database 2.0
Owner of data set	Plastics Europe
Copyright	Yes
Access and use restrictions	The data set can be used free of charge by anybody to perform LCA studies, to distribute it to third parties, to convert it to other formats, to develop own data sets etc. as long as the copyright and license conditions for the ELCD data sets and the ILCD format are met that can be accessed via http://lca.jrc.ec.europa.eu. Please note e.g. that reference must be given to the 'Owner of data set' and to the 'ELCD database' plus version number, when using the data set or parts thereof. Please note also, that any modifications/omissions of the data set results in invalidity of any existing 'Official approval of data set by producer/operator', that the impression must be avoided that this would still be a complete ELCD data set, and that the content of further fields has to be adjusted. For details see the aforementioned copyright and license conditions.

- Inputs and Outputs

nputs

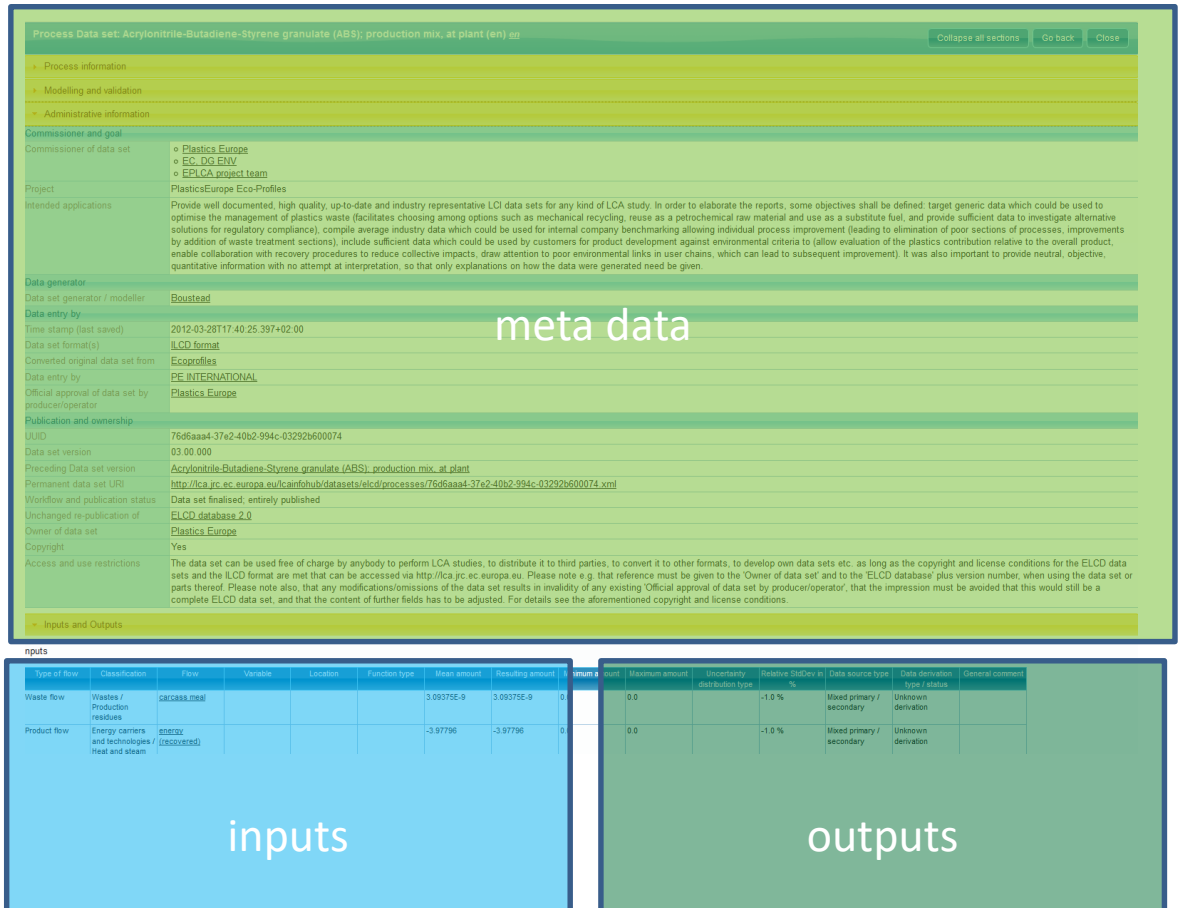
Type of flow	Classification	Flow	Variable	Location	Function type	Mean amount	Resulting amount	Minimum amount	Maximum amount	Uncertainty distribution type	Relative StdDev in %	Data source type	Data derivation type / status	General comment
Waste flow	Wastes / Production residues	carcass meal				3.09375E-9	3.09375E-9	0.0	0.0		-1.0 %	Mixed primary / secondary	Unknown derivation	
Product flow	Energy carriers and technologies / Heat and steam	energy (recovered)				-3.97796	-3.97796	0.0	0.0		-1.0 %	Mixed primary / secondary	Unknown derivation	

GreenDeLTa

Primary raw data acquisition

Background

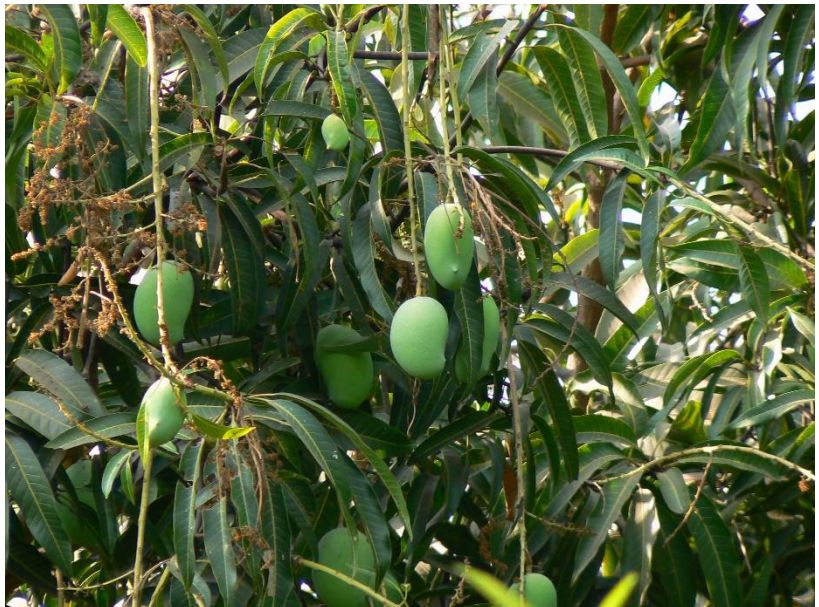
- How does a process data set in principle look like?



GreenDeLTa

Primary raw data acquisition
Mango production Argentina

- → What kind of information do we need to collect and store for a data set?
  - → Metadata?
  - → Inputs?
  - → Outputs?



Primary raw data acquisition
Mango production Argentina

→ Like this?

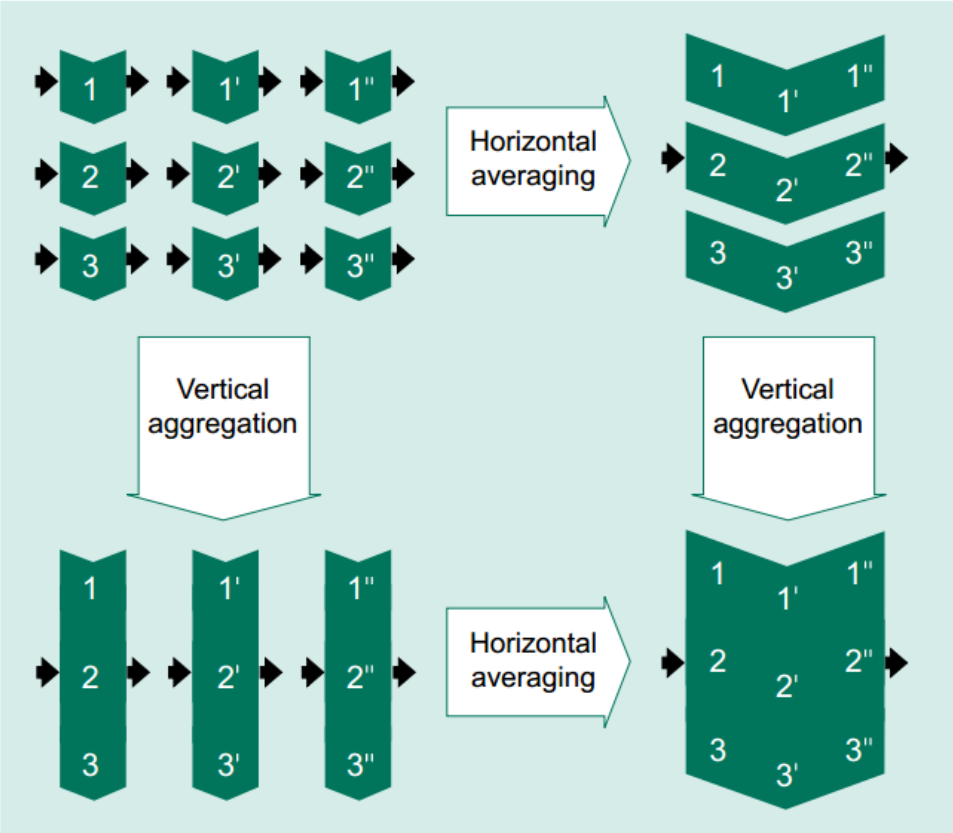
Metadata		
Author:	Huellas Ambientales lecture	
Year:	2018	
Region:	Argentina	
Data collection:	primary	
Input		
agricultural land use	0.5	m2/a
N-fertilizer	0.05	kg
manure	0.05	kg
pesticide xy	0.01	l
irrigation water, from grou	2	kg
biogenic CO2, from air	1	kg
Output		
Mango, fresh	1	kg
fertiliser, in ground	0.025	kg

Primary raw data acquisition
Protecting sensitive information

- Why is information collected for LCA potentially sensitive?
- Protecting sensitive information: approaches

Primary raw data acquisition
Protecting sensitive information

- Why is information collected for LCA potentially sensitive?
- Protecting sensitive information
  - Averages and aggregation



Broadbent, Cl., et al.: Aggregated data development, chapter 3 in UNEP Shonan Guidance Principles, UNEP 2012

GreenDelta

Primary raw data acquisition
Protecting sensitive information

- Why is information collected for LCA potentially sensitive?
- Protecting sensitive information
  - Averages and aggregation
  - Remove specificity
  - Is the information really sensitive?
- Protection by trusted “man in the middle”
- Self-Protection by data owner

Primary raw data acquisition
How to create representative data sets

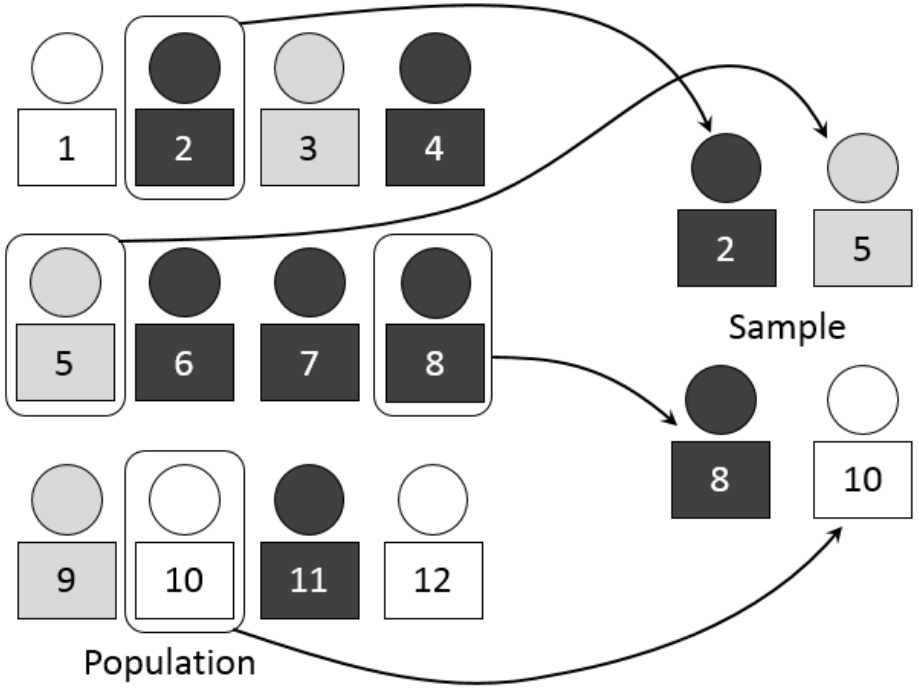
- What are representative datasets and why are they often needed?

Primary raw data acquisition
How to create representative data sets

- Common practice in LCA: market share
- This is not corresponding to scientific practice!
- Scientific: Statistical sampling

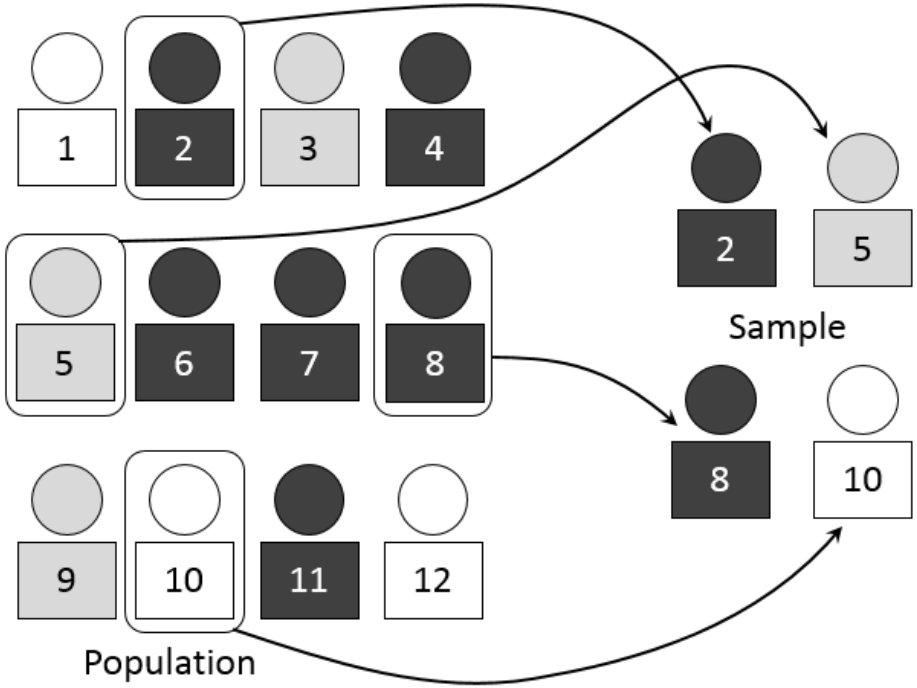
Primary raw data acquisition

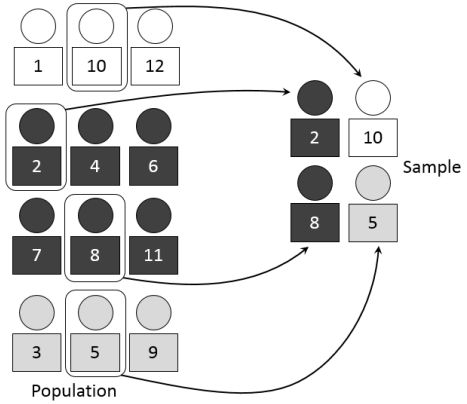
How to create representative data sets: statistical sampling



Primary raw data acquisition

How to create representative data sets: statistical sampling

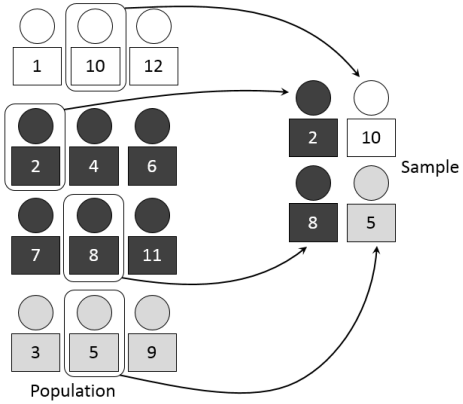


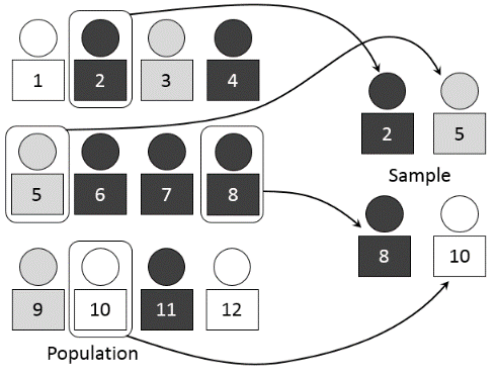


Primary raw data acquisition

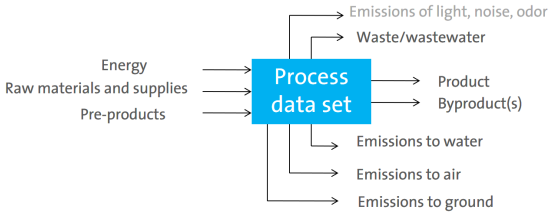
How to create representative data sets: statistical sampling

- Would you see any issues in applying this technique for LCA data?





Secondary data acquisition
Background

- 

- What are secondary sources that can be used here?

Secondary data acquisition
Secondary data

- Public, official statistics
- Wikipedia
- Company websites
- Journal papers
- Chemical engineering books
- Other LCA databases and datasets (!)
- …

Secondary data acquisition

Secondary data, example

Sri Lanka Energy Balance

Compiled by Sri Lanka Sustainable Energy Authority

Home | Electricity | Petroleum Products | Biomass | Energy Balance | Economic Indicators | Grid Emission | Conversion Factors | Coal

- Home
- Electricity Data
  - Overview
  - Electricity Generation
  - Own Use in Power Plants and Network Losses
  - Fuel Consumption in Power Plants
  - Electricity Sales by Sector
  - Capacities of Power Plants
  - Electricity Tariff
  - Customers, Employees and Finances
  - System Load Profile
- Petroleum Data
- Biomass Data
- Energy Balance
- Energy-Econ Indicators
- Grid Emission Factor
- Conversion Factors
- Coal

Gross Electricity Generation (GWh)

Select Year 2011 to 2015 Update Table Export to Excel

Links

Gross Generation of Thermal Power Stations
Gross Generation of Hydro Power Stations

CEB-Hydro (GWh)

	2011	2012	2013	2014	2015
Major Hydro	3,906.96	2,684.17	5,936.61	3,618.83	4,825.65
Small Hydro	65.72	42.55	73.49	30.89	78.76
Sub Total Hydro	3,972.67	2,726.72	6,010.10	3,649.72	4,904.41

CEB Non-Conventional (GWh)

	2011	2012	2013	2014	2015
CEB Wind	2.66	2.32	2.32	2.13	1.07
Sub Total Non Conventional	2.66	2.32	2.32	2.13	1.07

CEB-Thermal (GWh)

	2011	2012	2013	2014	2015
Steam - Fuel Oil	0.00	0.00	0.00	0.00	0.00
Steam - Coal	1,027.62	1,399.12	1,465.39	3,505.55	4,447.21
Steam - Diesel	10.49	4.61	3.98	19.40	9.97
Sub Total Steam	1,038.11	1,403.74	1,469.37	3,524.95	4,457.18
Diesel Engine - Residual Oil	903.37	921.55	559.81	647.10	271.93
Diesel Engine - Fuel Oil	0.00	0.00	111.48	95.69	87.94
Diesel Engine - Diesel Oil	14.19	9.42	27.26	9.31	22.48
Sub Total Diesel Engines	917.56	930.97	698.55	752.09	382.34
Gas Turbines - Diesel Oil	320.33	218.20	17.58	241.88	25.07
Gas Turbines - Naptha	0.00	0.00	0.00	0.00	0.00
Sub Total Gas Turbines	320.33	218.20	17.58	241.88	25.07
Combined Cycle - Diesel Oil	95.90	550.66	221.75	284.64	119.54
Combined Cycle - Naptha	159.78	329.05	388.53	465.54	540.26
Sub Total Combined Cycle	255.68	879.71	610.27	750.18	659.80

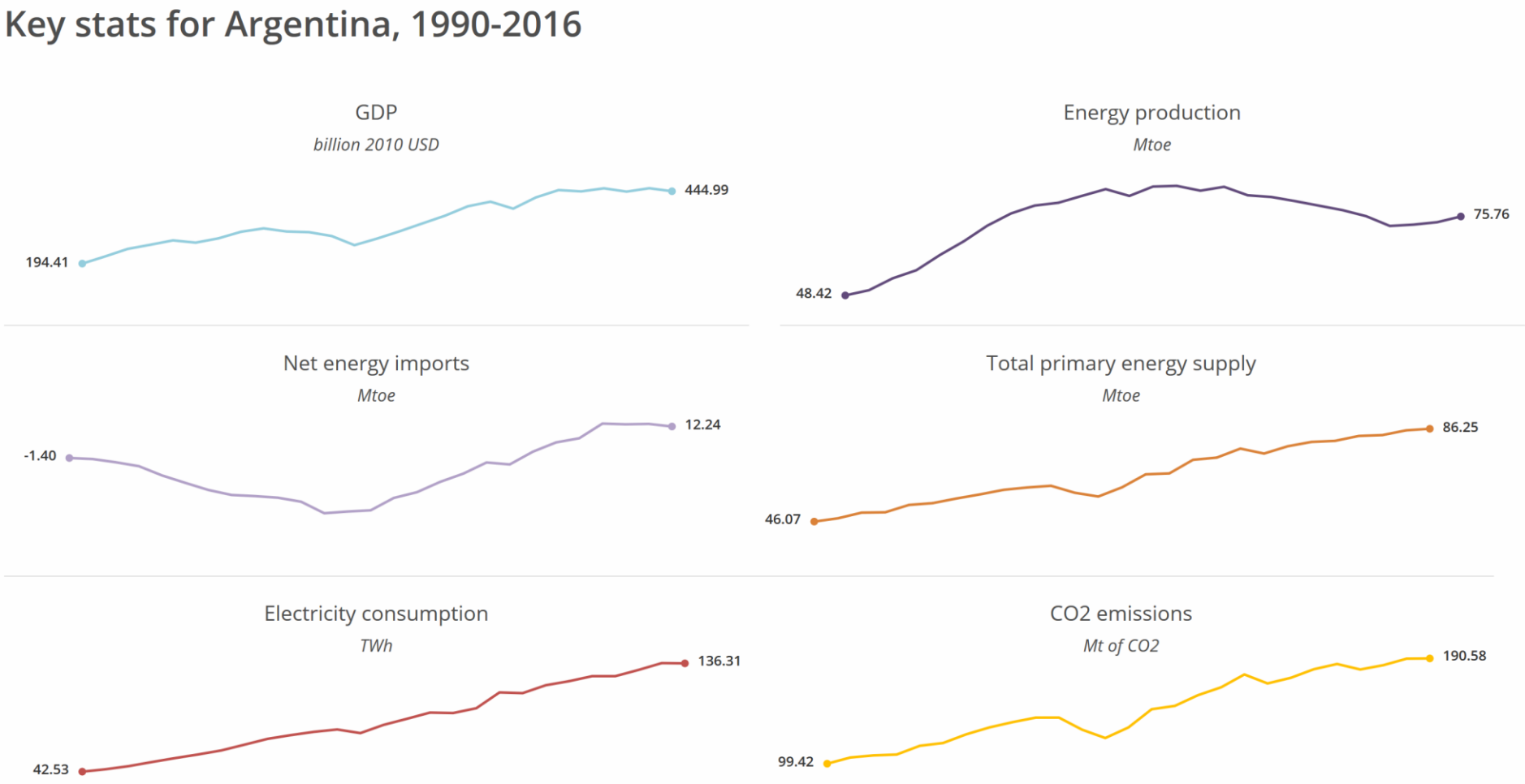
CEB - Sub Total (GWh)

	2011	2012	2013	2014	2015	View Diagrams

http://www.info.energy.gov.lk/

GreenDeLTa

Secondary data acquisition
Secondary data, example

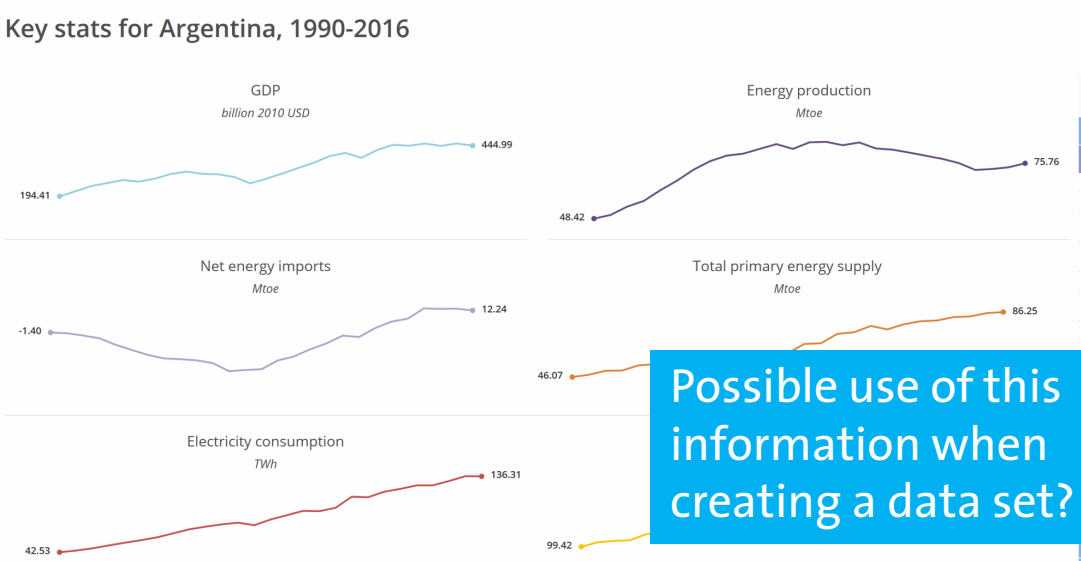


https://www.iea.org/countries/Argentina/

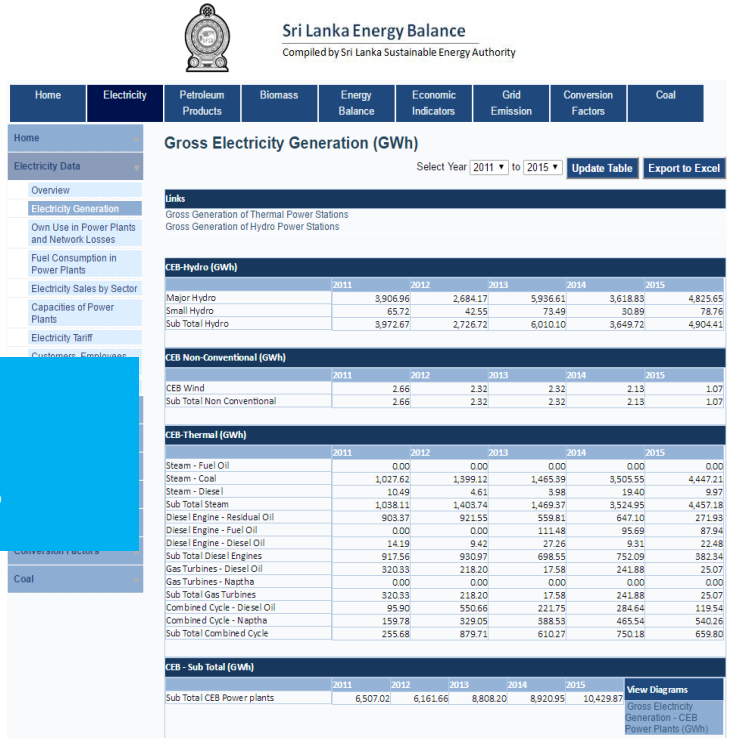
GreenDelta

Secondary data acquisition

Secondary data, example



Possible use of this information when creating a data set?



GreenDeLTa

Secondary data acquisition
Secondary data – challenges, issues

- Quite often, data does not fully fit to what you need
  - Incomplete
    - Product missing
    - Only some emissions reported, or only composition of products
  - Not fully fitting (slightly different product, location, time than needed)
- Source reliability

Secondary data acquisition
Secondary data – challenges, issues

- Quite often, data does not fully fit to what you need
  - Incomplete
    - Product missing
    - Only some emissions reported, or only composition of products
  - Not fully fitting (slightly different product, location, time than needed)
- Source reliability

→ Need attention and care, but great potential. Typically, result are "patchwork datasets"

Live presentation of datasets in a database, LCA model (if possible)

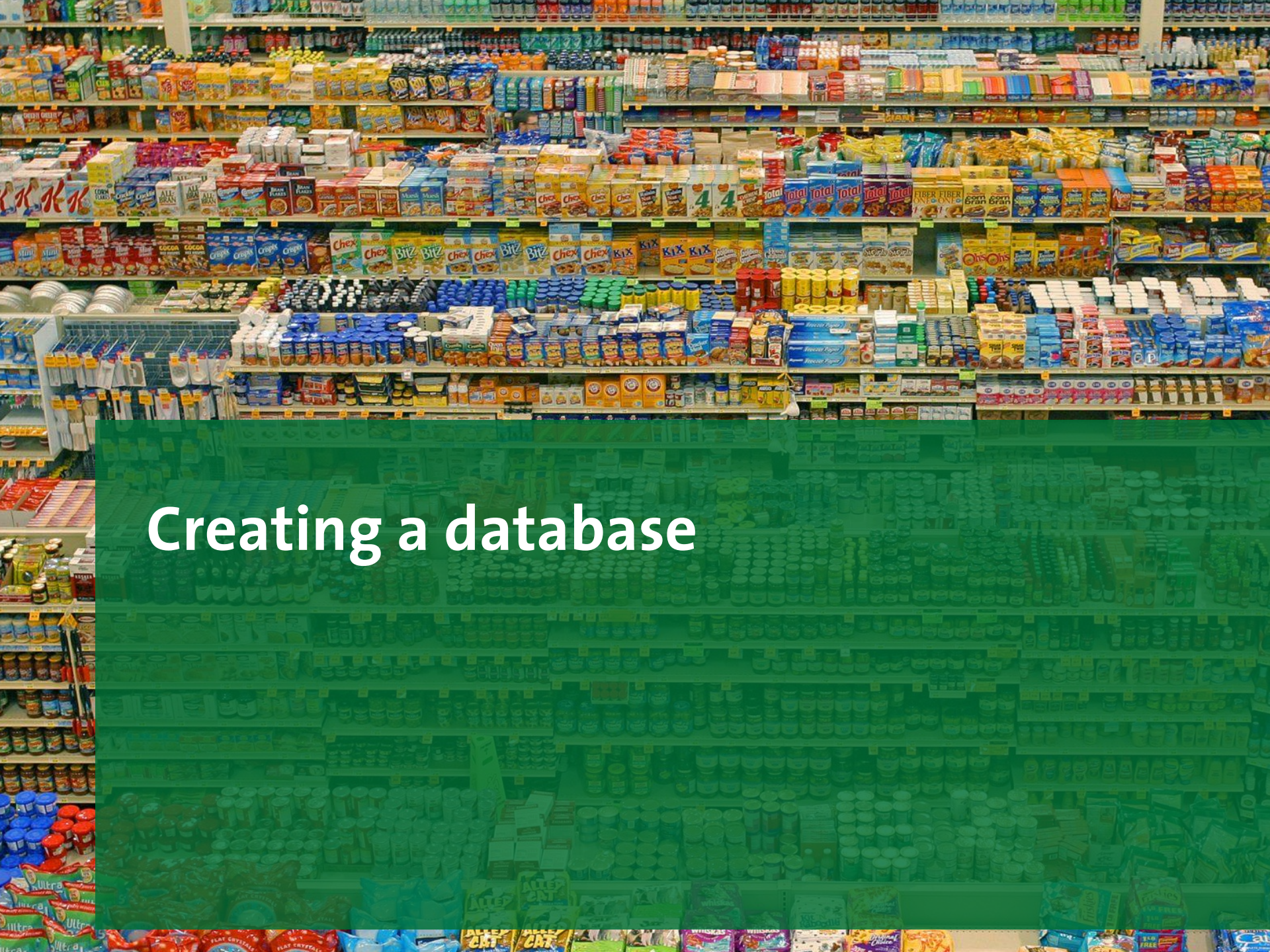
Data set creation, final thoughts

Issues when creating many datasets, in a group?

Data set creation, final thoughts

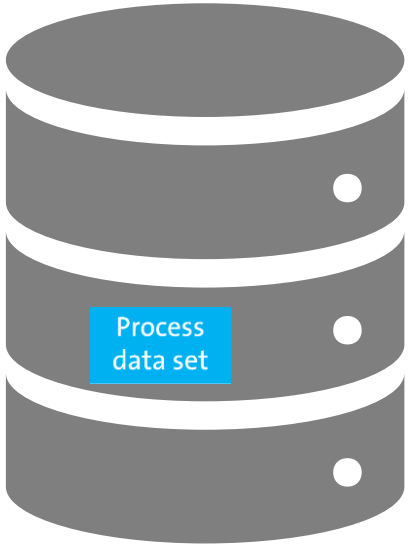
Issues when creating many datasets, in a group?

- consistency of flows
- consistency of modeling
- timing (creating the same process twice, building on other processes that have already been created)
- updating datasets in a group
- …



Creating a database

Datasets in databases



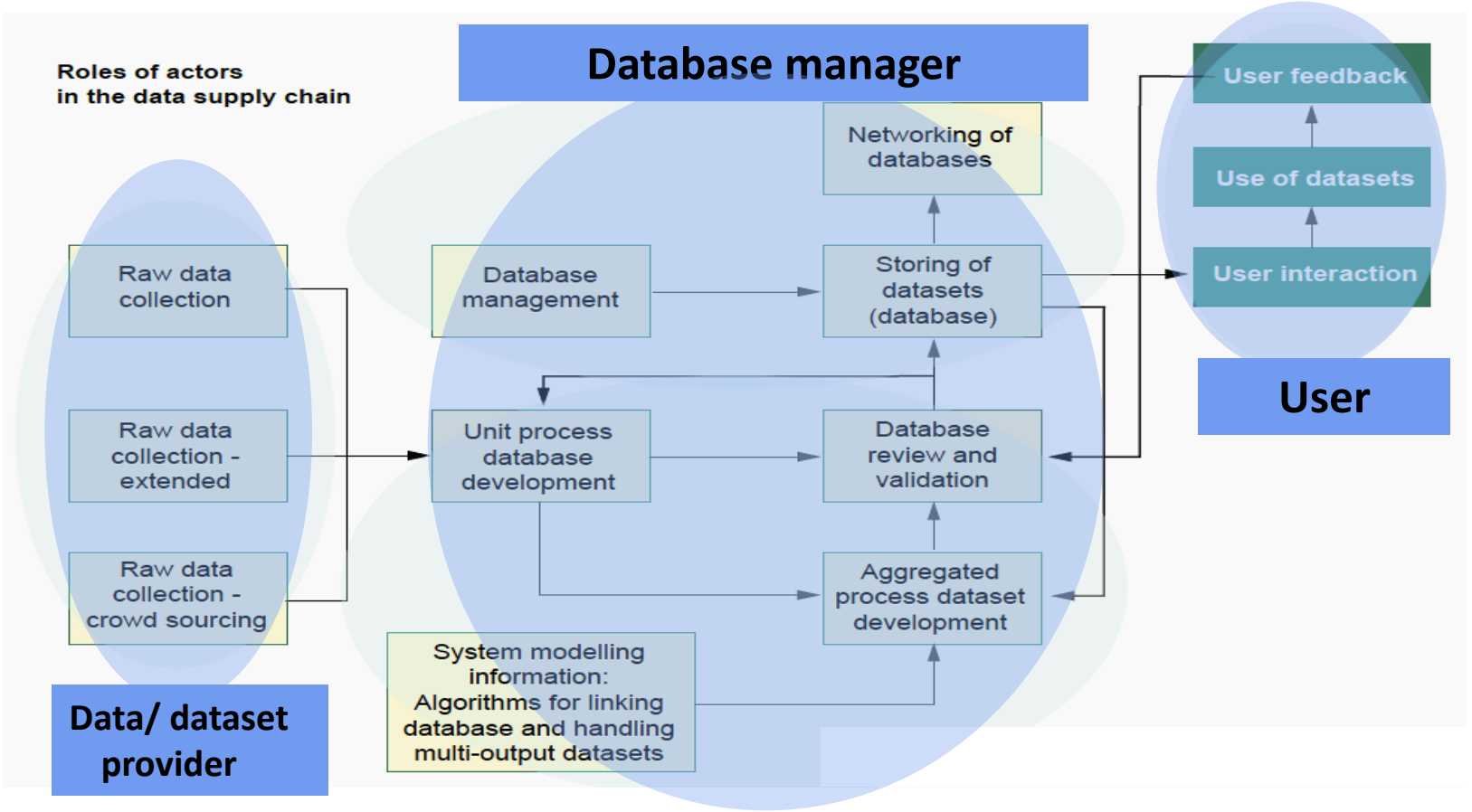
Requirements on Datasets in databases

- Comprehensive input/output flows for datasets
- Consistent modeling approach and flow nomenclature
- Complete and consistent dataset documentation

LCA database consistency

- often, the database has capability beyond storing data, e.g. data collection and submittal support, review, aggregated process creation, ...
- consistency in modeling approach allows combination of the datasets in a Life Cycle model
- if several databases are consistent, they can also be combined

Datasets in databases
The whole picture:
data provider, DB manager, user



Sonnemann, G. Vigon, B. et al.: The Context for Global Guidance Principles for Life Cycle Inventories, chapter 1 in UNEP Shonan Guidance Principles, UNEP 2012

Datasets in databases
Format of the database

- In the past, long discussions in various emerging databases, e.g. Brazil, US, ... about the “right” format
- Discussions are mainly about the exchange format (-> some few, from ecoinvent, JRC/EC, GreenDelta)
- For the actual storage, the format is not so important but DB should be able to “deliver” in various broadly used formats
- Format in use for storage:
  - LCDN / EC: ILCD (only!)
  - USDA: openLCA
  - Ecoinvent: unknown

Datasets in databases
Version control

- Typically, database users expect stability in their results
- ..but they also expect updates, bug fixes, extensions of the database

Datasets in databases
Version control

- Typically, database users expect stability in their results
- ..but they also expect updates, bug fixes, extensions of the database
- Common practice:
  - Release versions
    - Ecoinvent: major, smaller release (3.5)
    - GaBi: Annual releases with mid-year update
  - Dataset versioning (ILCD hubs e.g.): each dataset has a version number and can be chosen accordingly
  - Collaboration server: push and commit dates and versions, with diff files, most sophisticated

Datasets in databases
Review and Quality Assurance

- Review “items” (what is reviewed, how)
- Reviewers
- Review workflow and organisation

Datasets in databases
Review and QA

- Review “items” (what is reviewed, how)
- Reviewers
- Review workflow and organization
- On which type of processes should the review be performed (unit process, aggregated process?)

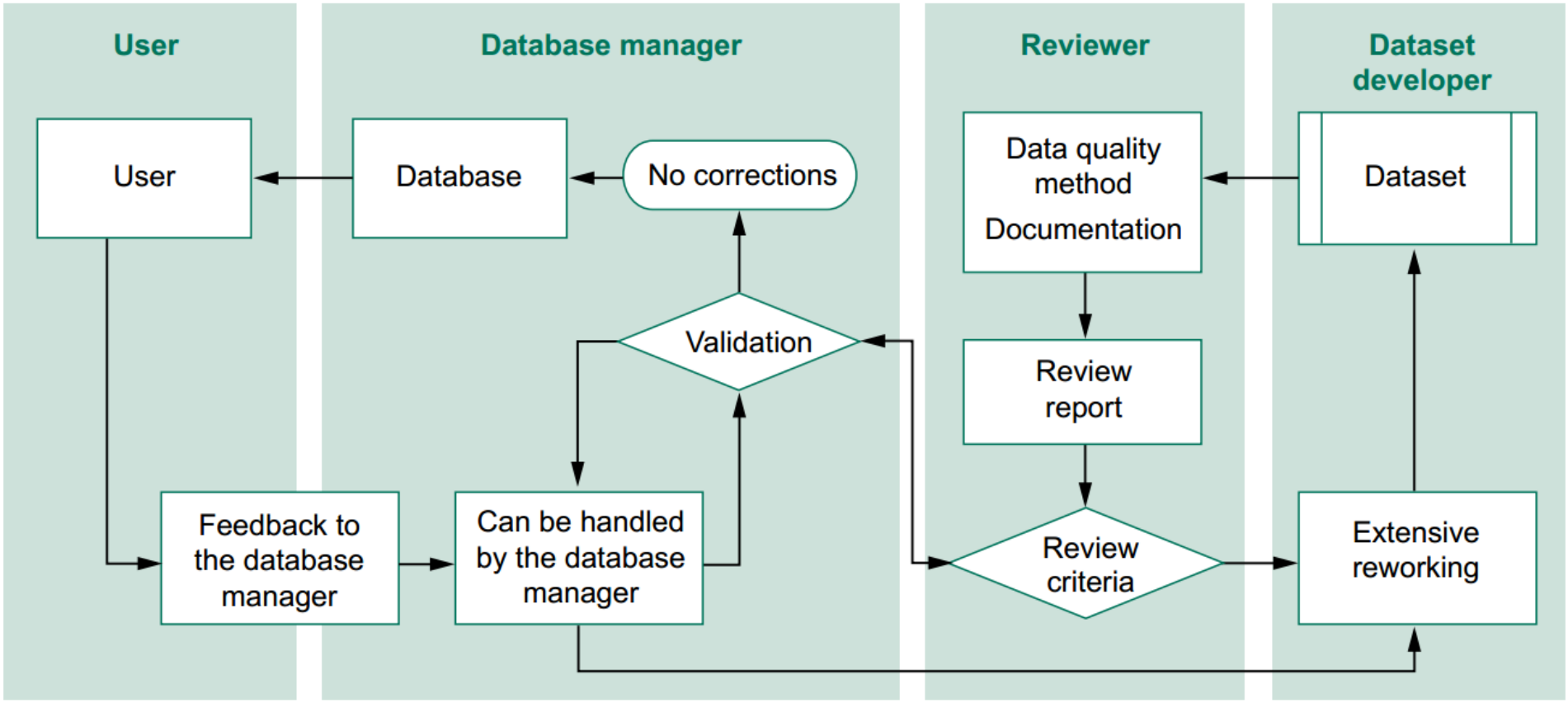
Datasets in databases

Review criteria

	Values	Scope		
		data set	flows/ exchanges	any other data set fields
Goal and scope completeness				
Reference time	yes/no	x		
Reference geography	yes/no	x		
Reference technology	yes/no	x		
Reference model completeness	yes/no	x		
Reference sample completeness	yes/no	x		
Sample approach (scientific, or expert-based)	scientic / expert	x		
Supported LCIA methods with version number	text	x		
Conformance				
Time related conformance	1..5		x	x
Geographical conformance	1..5		x	x
Technological conformance	1..5		x	x
Model completeness conformance, flows and documentation	1..5	x		
Sample conformance, correctness and reliability				
Sample conformance	1..5	x		
Accuracy of the provided information	1..5	x		
Precision of the provided information	1..4		x	
Reliability of the provided information	1..5		x	x
Consistency of the provided information	1..5		x	x
Materiality				
Mass- and energy balance in line with goal and scope	1..5	x		
LCIA results in line with goal and scope	1..5	x		
Order of 5 main drivers for main LCIA results in line with goal and scope	1..5	x		
Procedural and meta-information				
# of reviewers and their relation to data provider	1..5	x		
Data access	1, 3, 5	x		

Ciroth, A., Foster, Chr., Hildenbrand, J., Zamagni, A.:Life Cycle Inventory Dataset Review Criteria Development, final report 2016

Datasets in databases
Review procedure, example



Inaba, A., et al.: Data Documentation, review, and management, chapter 4 in UNEP Shonan Guidance Principles, UNEP 2012

Datasets in databases
Standards for the review of datasets?

- ISO 14040 is well known as main standard for LCA
- there is also a standard on review, which does not cover datasets however (ISO 14071 TS -> “LCA study”)
- in PEF of the European Commission, and other schemes, there is guidance for performing the review

→ a dedicated standard, and overall accepted approach, for datasets review does not exist yet (!)

Datasets in databases
Review, software support

There exist tools to support the review procedure, including

- EcoEditor, ecoinvent (fully integrated)
- ILCD validation tool (ILCD, for nomenclature, documentation completeness, ... - not content, not review procedure)
- LCA Collaboration Server (fully integrated), by GreenDelta, e.g. USDA

(maybe more)

Creating a national LCA database

(i.e., a database that is national reference, often provided for free, and maintained by public institutions)

- LCA databases first existed for several EU countries, US, Japan, Australia
- In other regions, processes are often very different → need for additional databases
- LCA data is like infrastructure -> free access to LCA data is a public service
- there is often need for an agreed value of e.g. electricity grid Greenhouse Gas emissions
- creating the database fosters national industry and science

Creating a national LCA database

Quite some national databases are now created or have been created since some time:

- Chile
- Brazil
- Malaysia
- Thailand
- India
- Sri Lanka
- …

This is a chance for local LCA communities, to bring in innovation, and to benefit from existing experiences at the same time

When creating a national LCA database, many decisions need to be taken

business model (long time support?)
which industry sectors, products to include
start and development, scale up procedure
data collection procedures
data exchange format
review system
nomenclature of flows and process datasets
modeling principles
data quality assurance in general
persons in charge
IT infrastructure
supported LCIA methods
connection to LCA software
connection with other existing LCA databases
unit processes, system processes, or even LCIA results only in the database
...

When creating a national LCA database, be prepared for stakeholder discussions

- LCA software providers
- LCA consultants and researchers
- national industry

→ LCA data is a market, and a national LCA database can impact this market (LCA software providers, other databases)

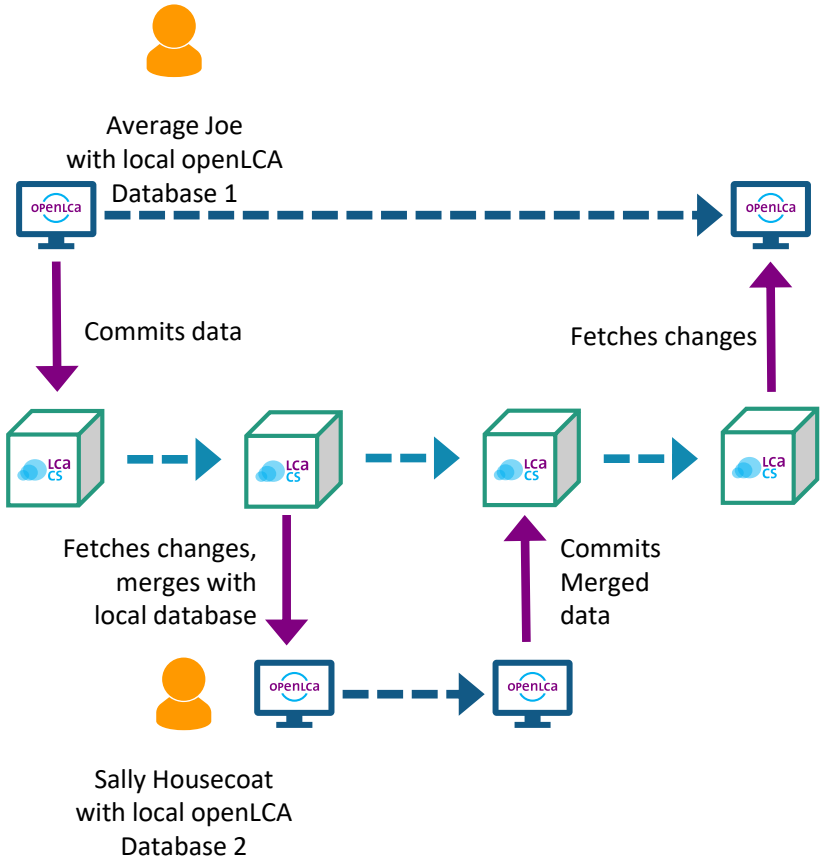
→ results from the national database are obviously interesting for national industry


LCA Collaboration Server

The LCA Collaboration Server

Introduction

- Server application that complements openLCA
- Facilitates the exchange of LCA datasets between users who work in a distributed team
- Synchronisation of databases, tracking of changes, comparison of databases
- The LCA CS empowers users to conduct distributed and simultaneous collaborative LCA modelling



The LCA Collaboration Server

local database, openLCA

Local database with projects, product systems, processes, flows, indicators and parameters as well as background data.

online repository, LCA Collaboration Server

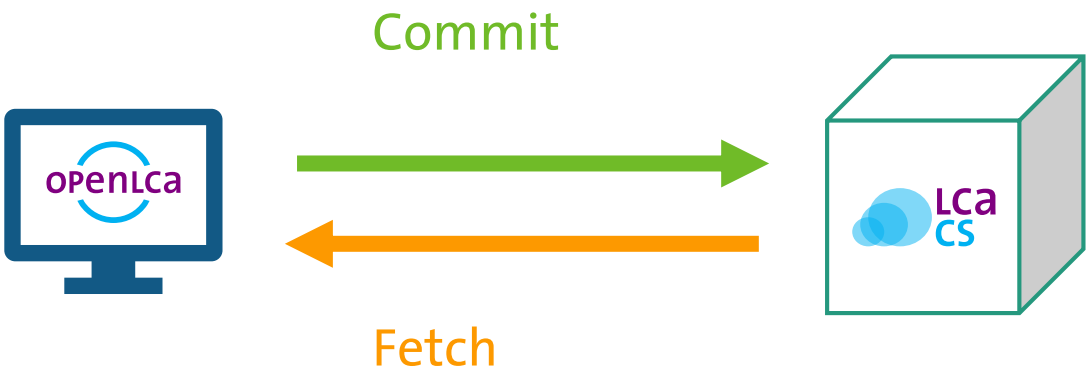
A repository is the equivalent to a database in openLCA and contains the same elements. A repository mirrors the local database (of the users connected to it).

You can have several repositories in one LCA CS.

The LCA Collaboration Server

2 Important terms

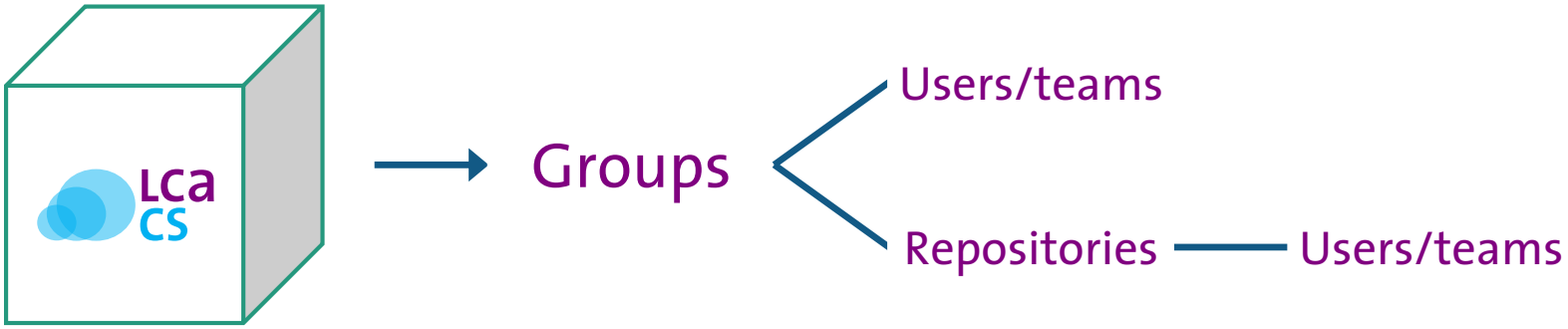
- **Commit:** Committing data means *pushing* data from a local openLCA database to an LCA CS repository
- **Fetch:** Fetching data means *downloading* data from an LCA CS repository to a local openLCA database



The LCA Collaboration Server

Structure

- For one LCA CS, you can define one or several groups, with several teams and / or users per group,
- ..and with one or several repositories, linked to users and teams



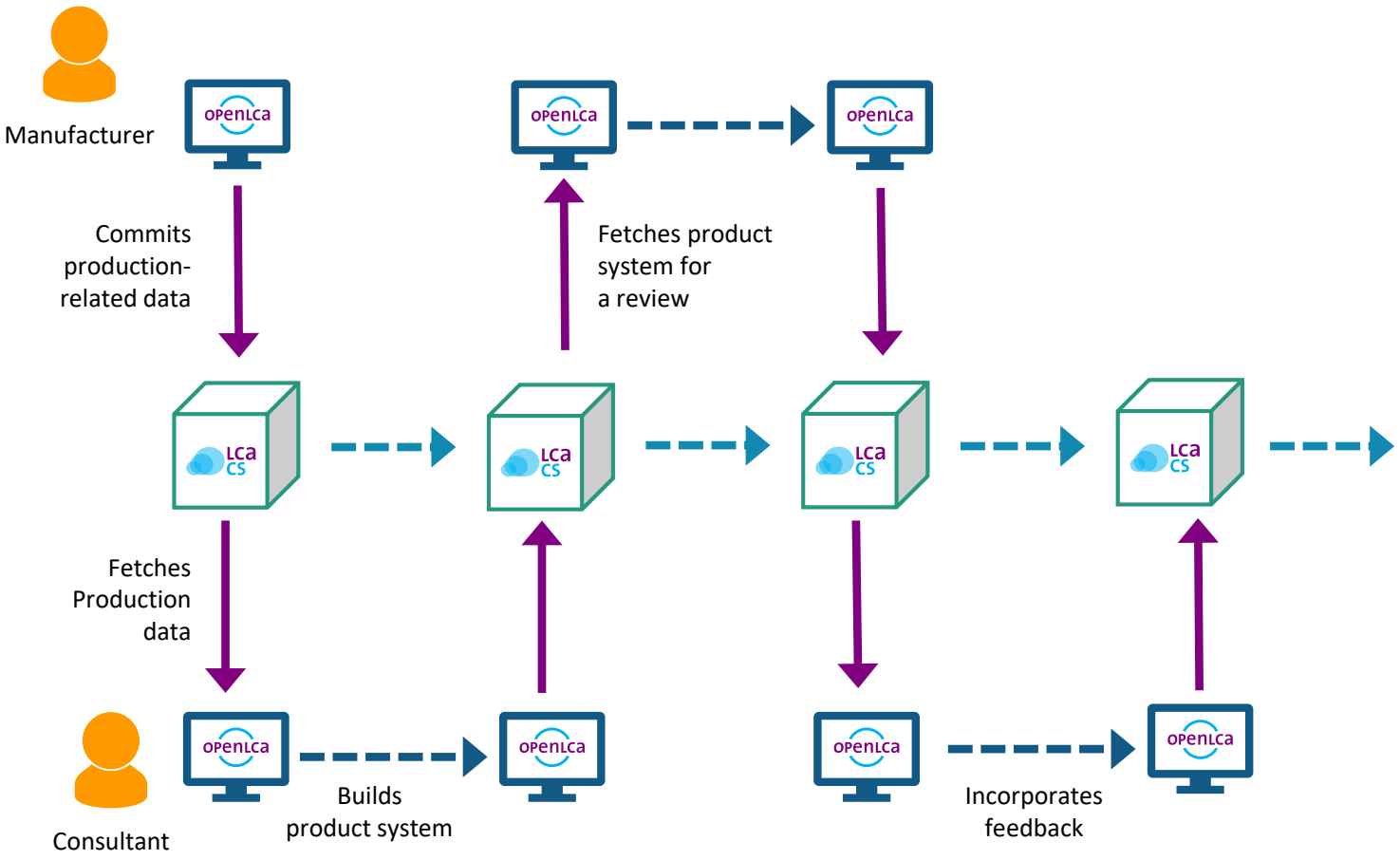
The LCA Collaboration Server

Group and user roles

Rights	Reader	Contributor	Reviewer	Editor	Owner
Read repositories and fetch contents	X	X	X	X	X
Commit data to repositories		X	X	X	X
Comment specific fields of data sets			X	X	X
Review comments			X	X	X
Manage comments				X	X
Manage reviews				X	X
Create repositories					X
Edit repository members					X
Adjust settings					X
Move repositories					X
Delete repositories					X

The LCA Collaboration Server: Use Cases

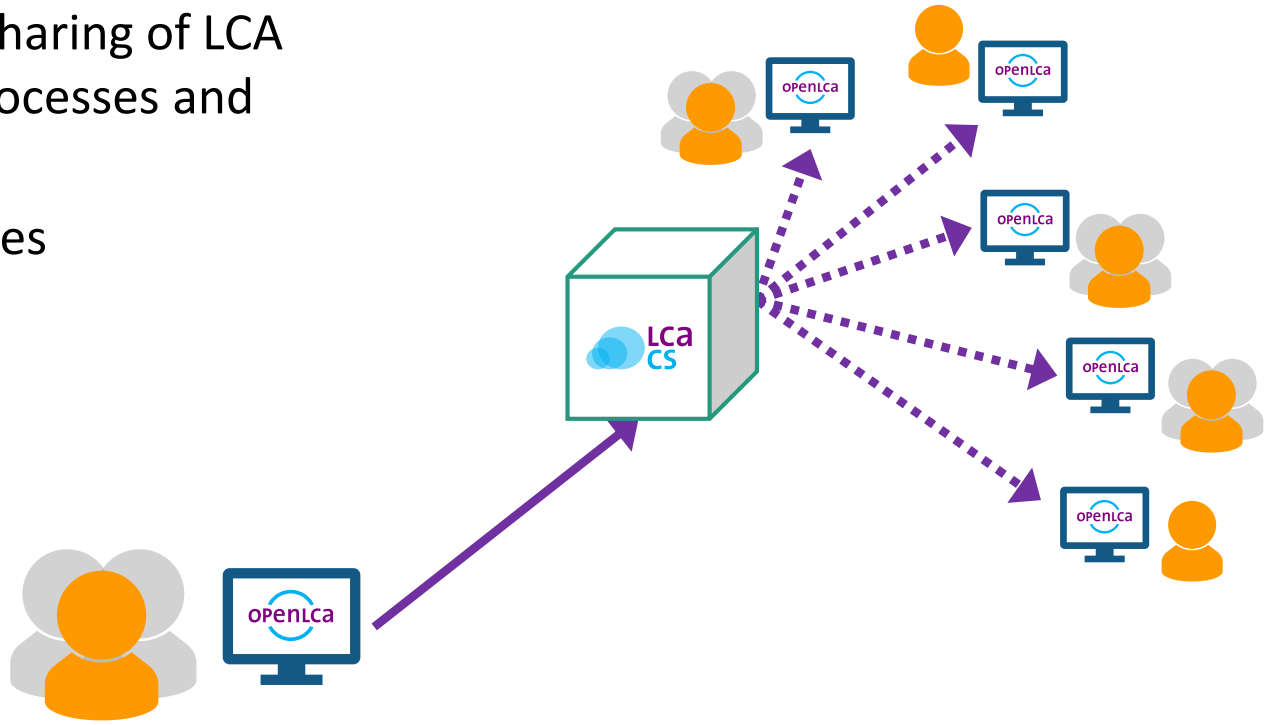
Iterative co-development of datasets and LCA models, data collection



The LCA Collaboration Server: Use Cases

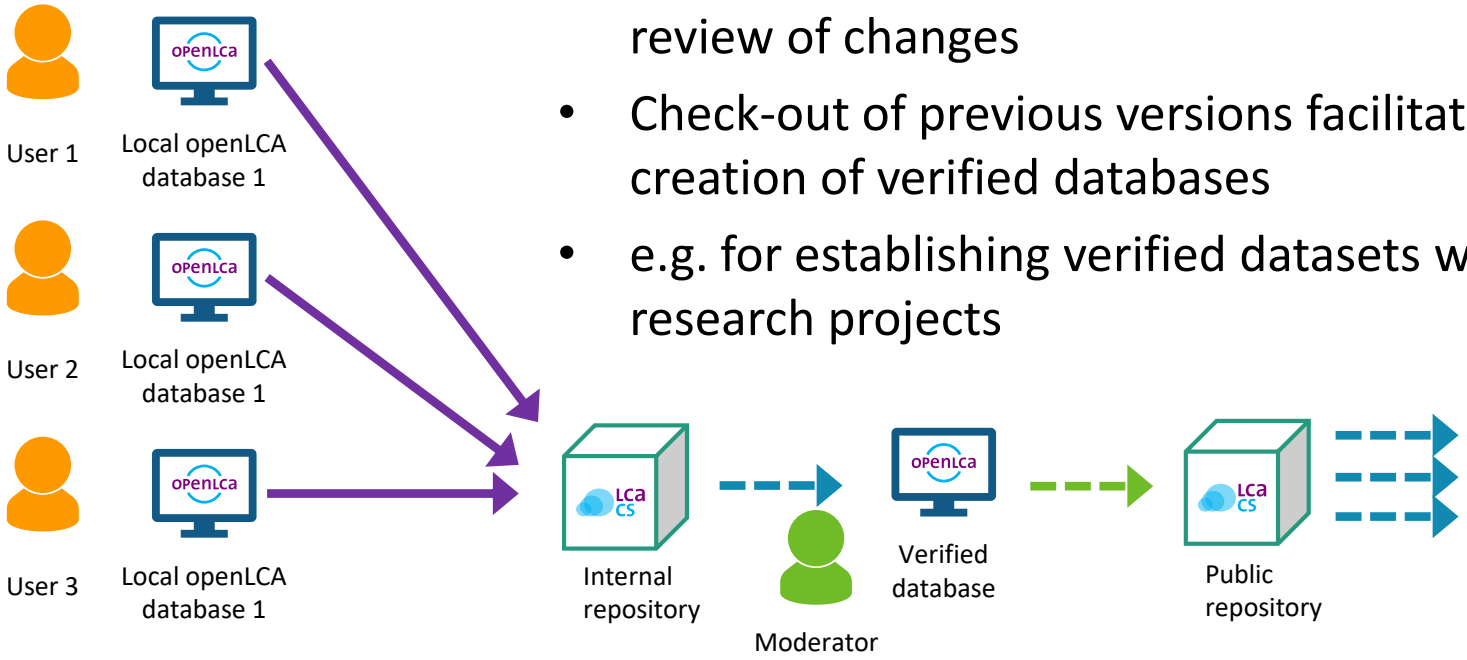
"Publication": Straightforward sharing of LCA models, flows, processes and entire databases

- Straightforward sharing of LCA models, flows, processes and entire databases
- Tracking of changes
- Comments



The LCA Collaboration Server: Use Cases

Building and managing verified public LCA repositories

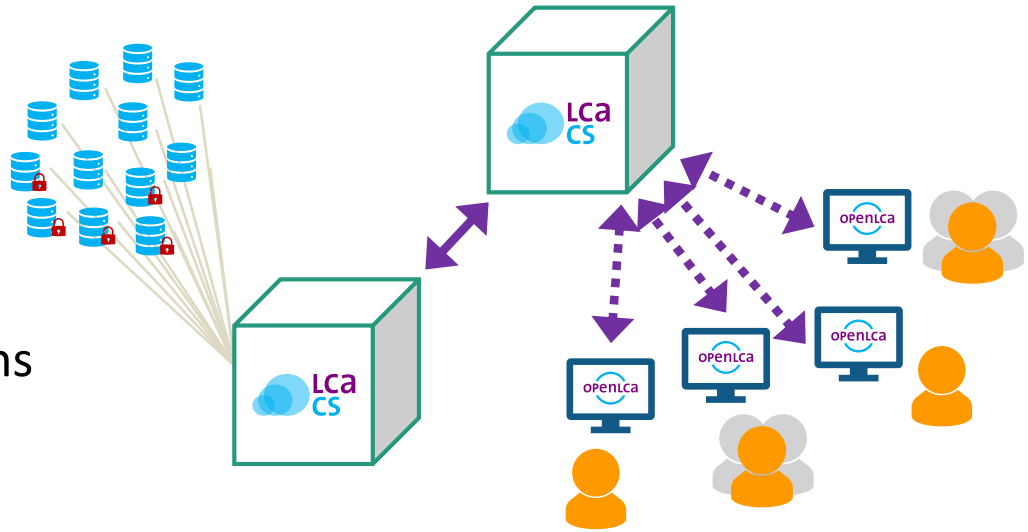


- Tracking feature and versioning allows simple review of changes
- Check-out of previous versions facilitates the creation of verified databases
- e.g. for establishing verified datasets within research projects

The LCA Collaboration Server: Use Cases

Managing and distributing reference data

- Many projects aim at establishing reference data for LCA studies that conform with a specific industrial or environmental standard
- Protect specific datasets whilst accepting contributions from other users



The LCA Collaboration Server

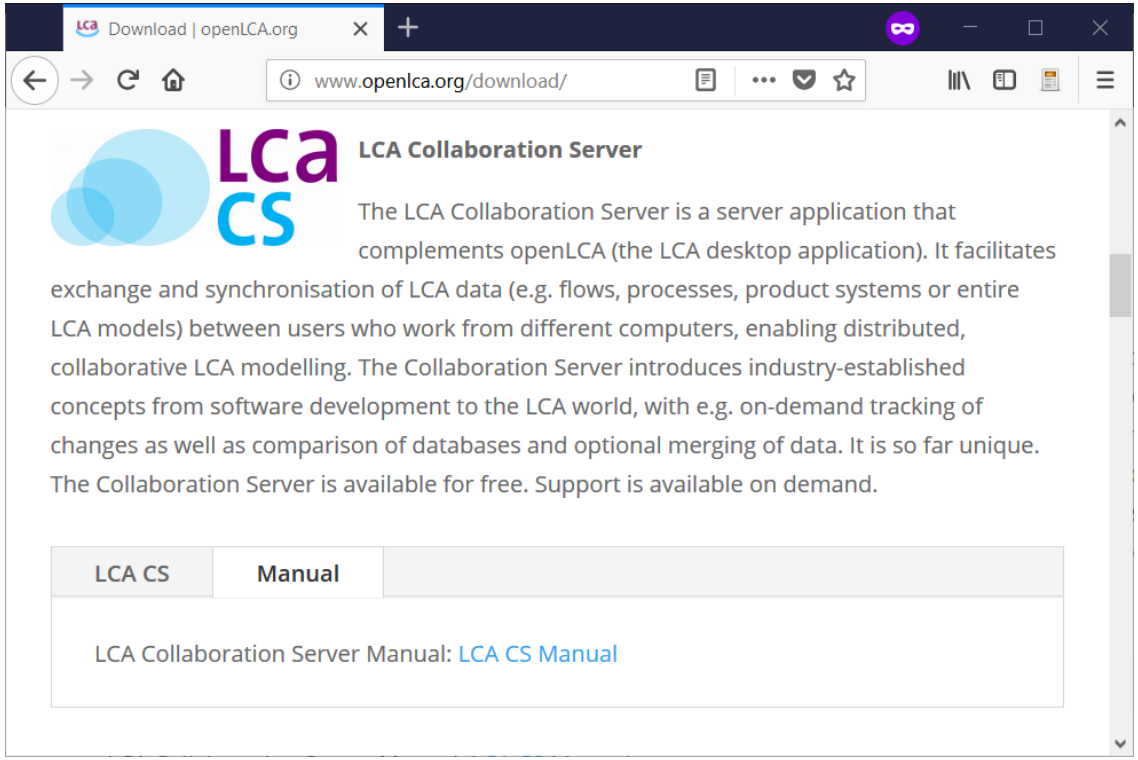
Download

Download the LCA CS and the LCA CS User manual via

http://www.openlca.org

Install on Webserver

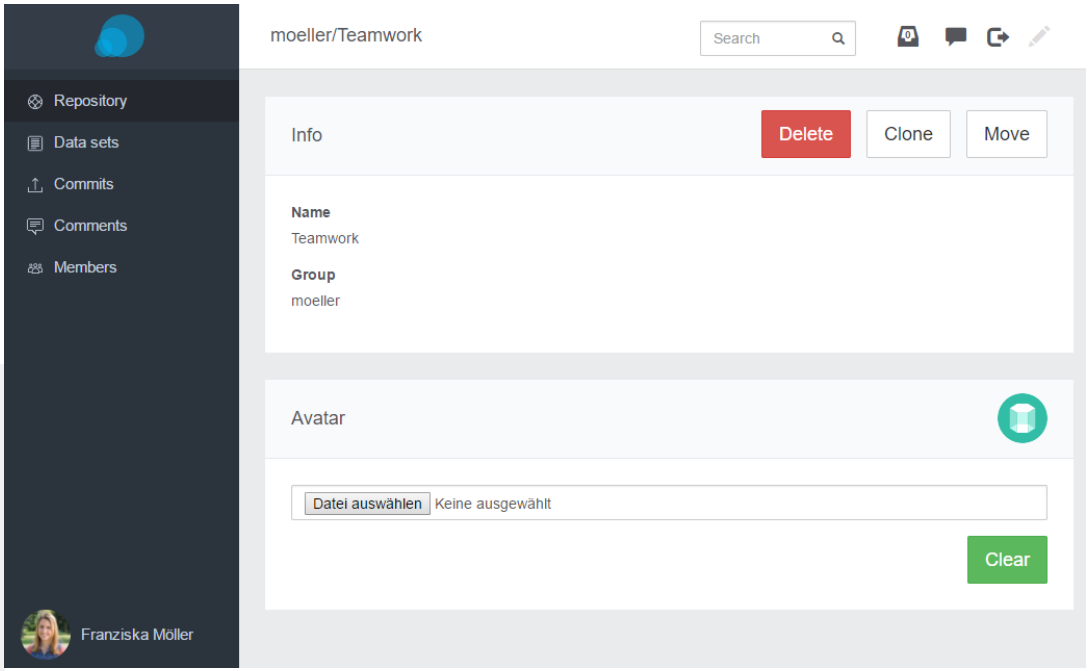
(GreenDelta-hosted solution, and support, available)

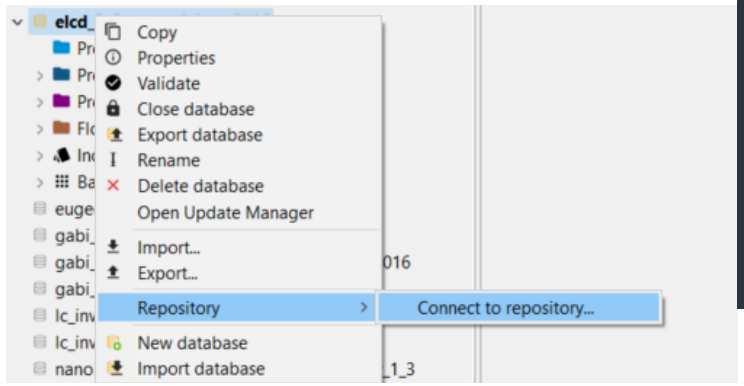


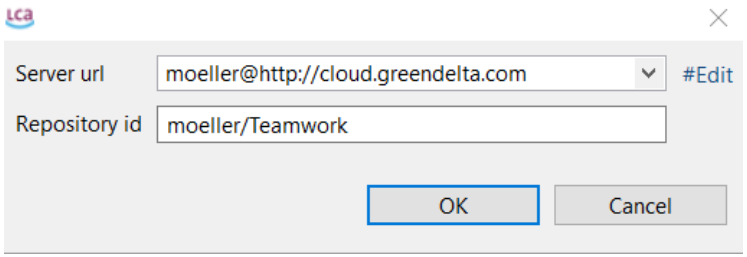
Exemplary use case – work flow

- A repository is created and User 1 and User 2 are members

- User 1 connects local db to repository

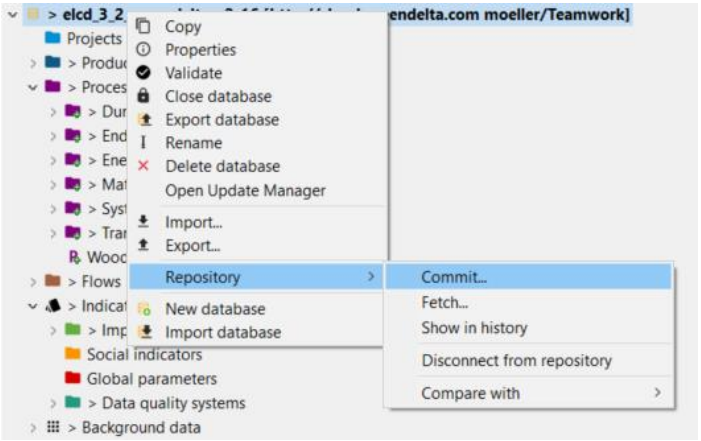


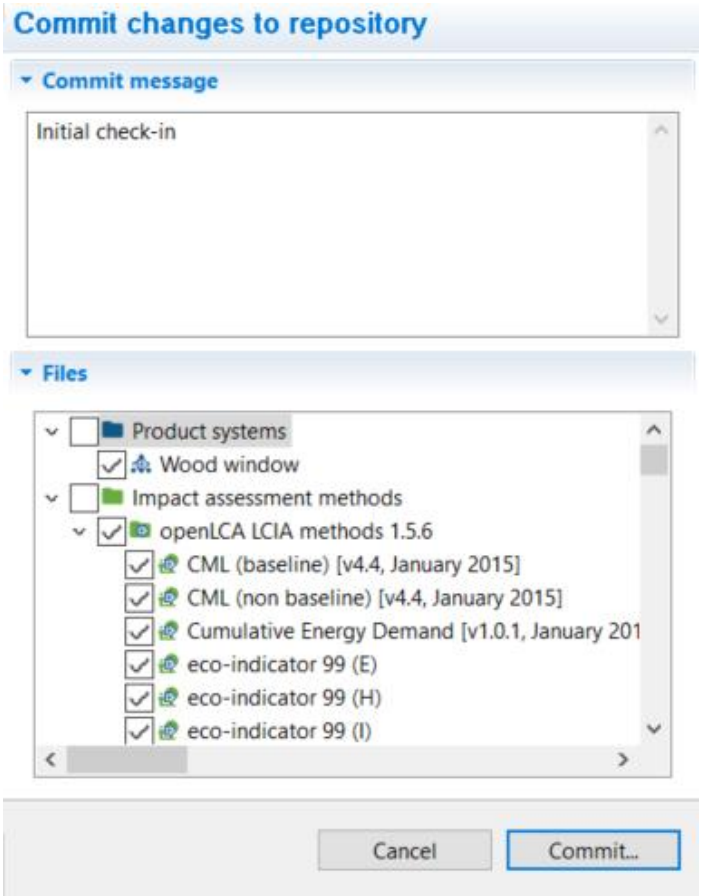




Exemplary use case – work flow

- User 1 commits data

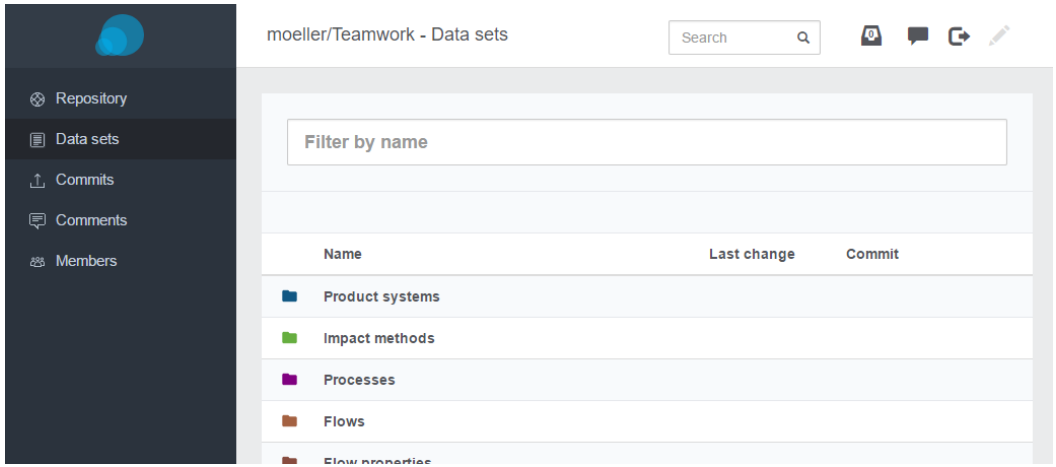


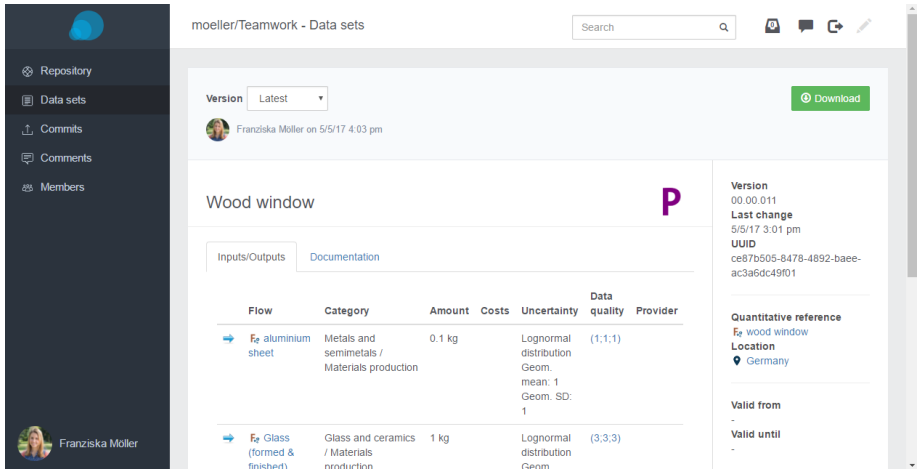


GreenDelta

Exemplary use case – work flow

- The data appears on the collaboration server

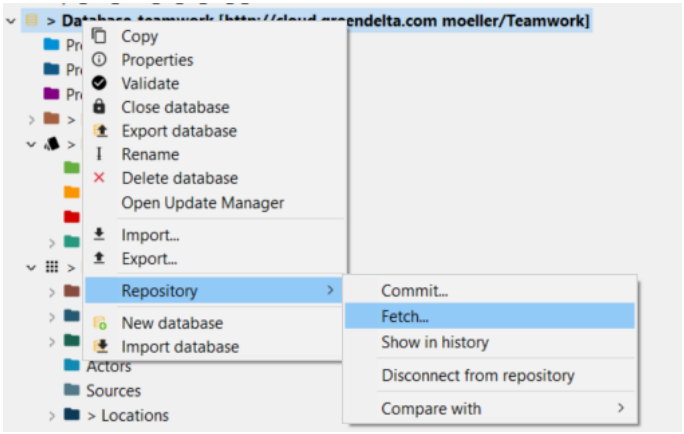


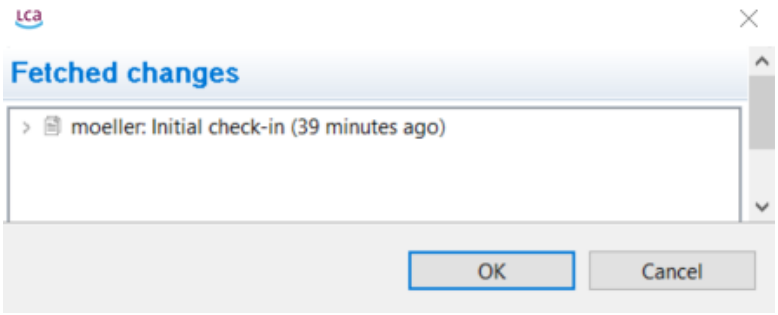


GreenDeLTa

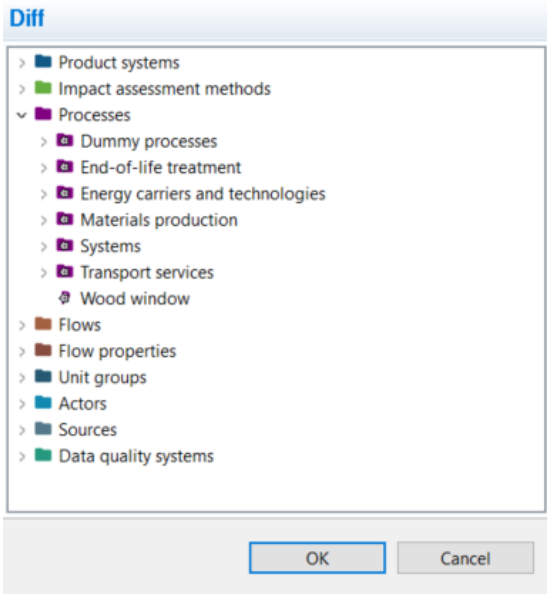
Exemplary use case – work flow

- User 2 connects to the same repository and fetches data





- Summary of differences to local database appears



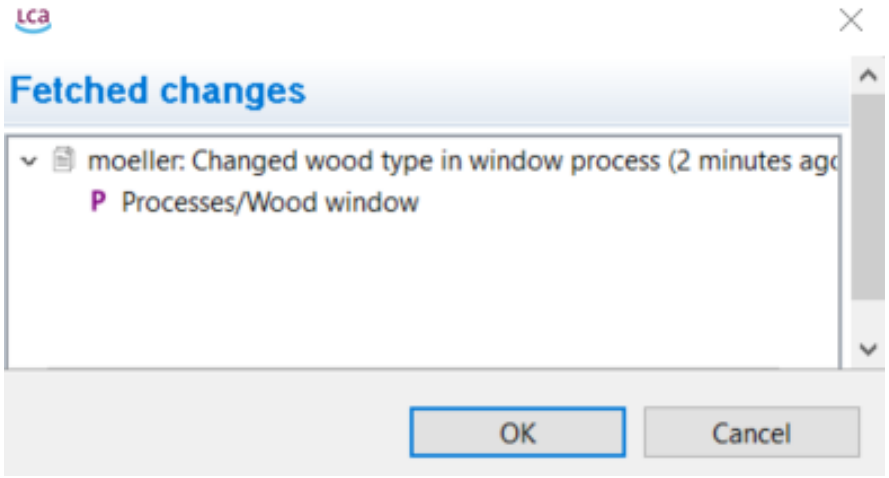
GreenDelta

Exemplary use case – work flow

- User 2 makes change in local data and commits again to repository

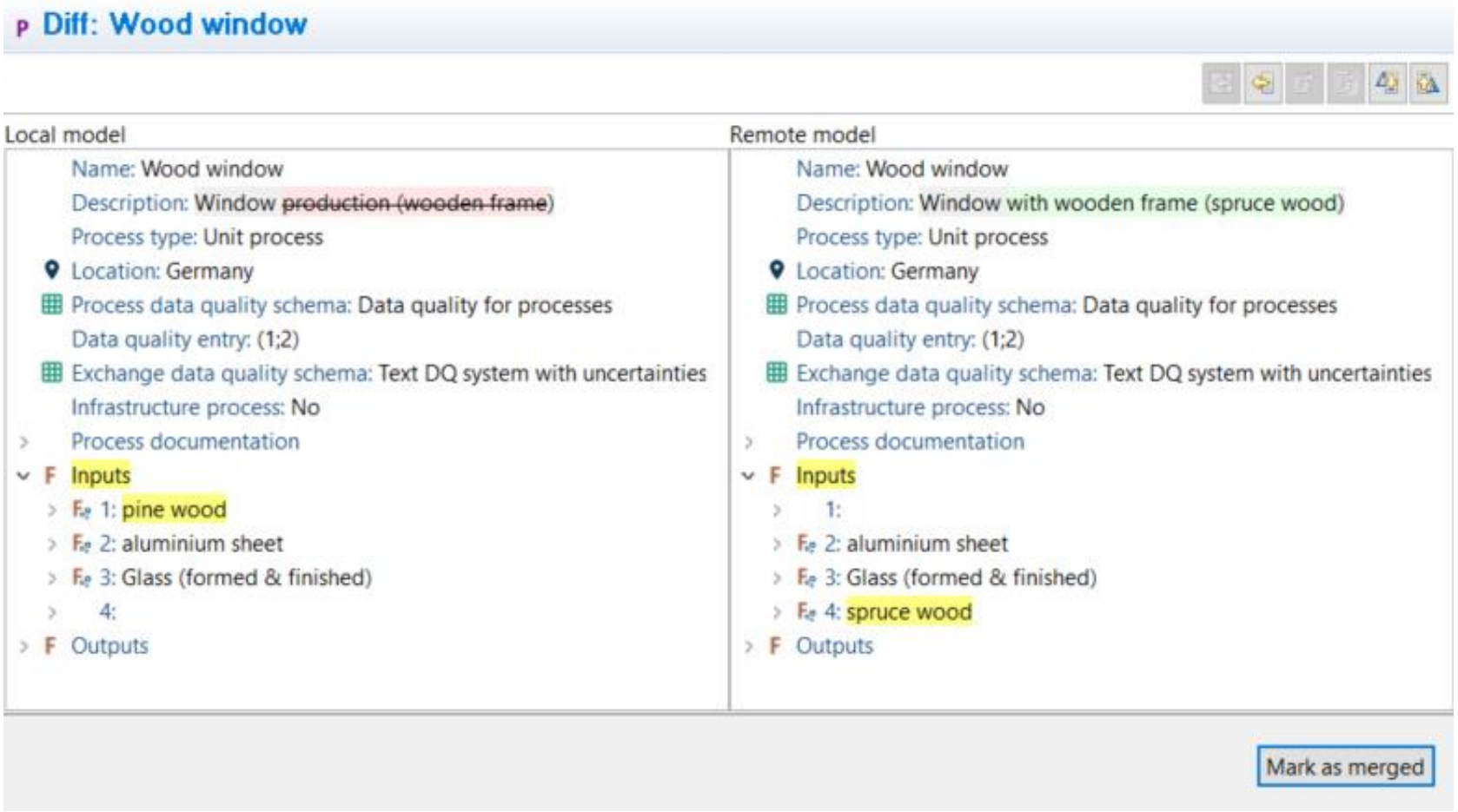
- User 1 makes change in local data and wants to commit

- User 1 fetches changes



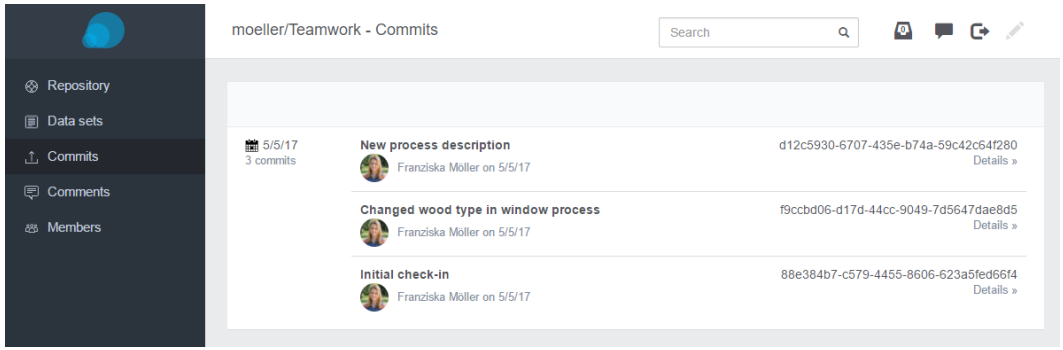
Exemplary use case – work flow

- Diff tool shows the differences in the data

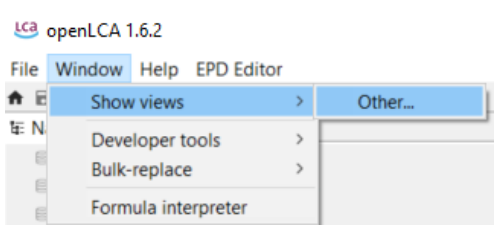


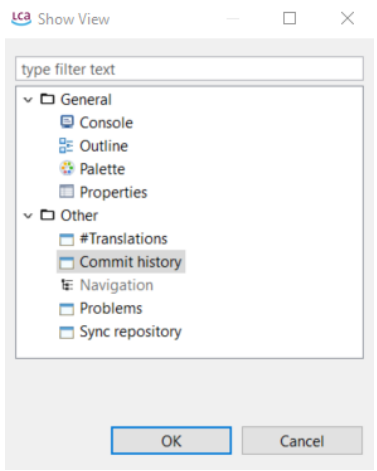
Exemplary use case – work flow

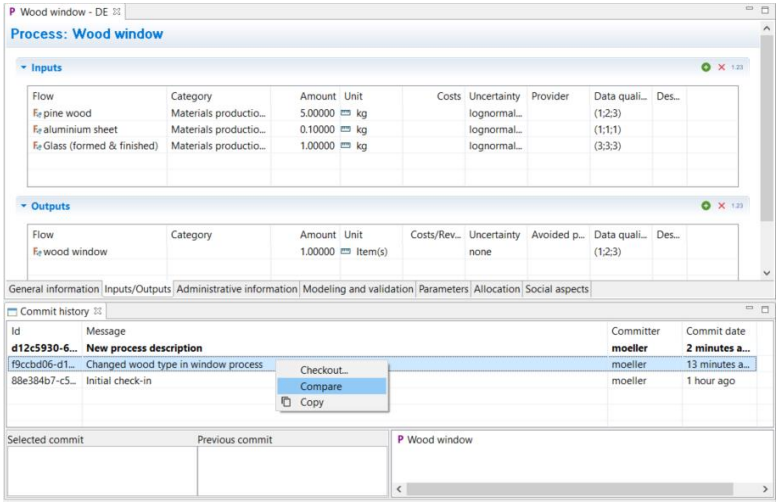
- Commit history in the web app



- Commit history in openLCA







GreenDeLTa

LCA Collaboration Server, status

… a central component of the US agencies' LCA toolbox[1]

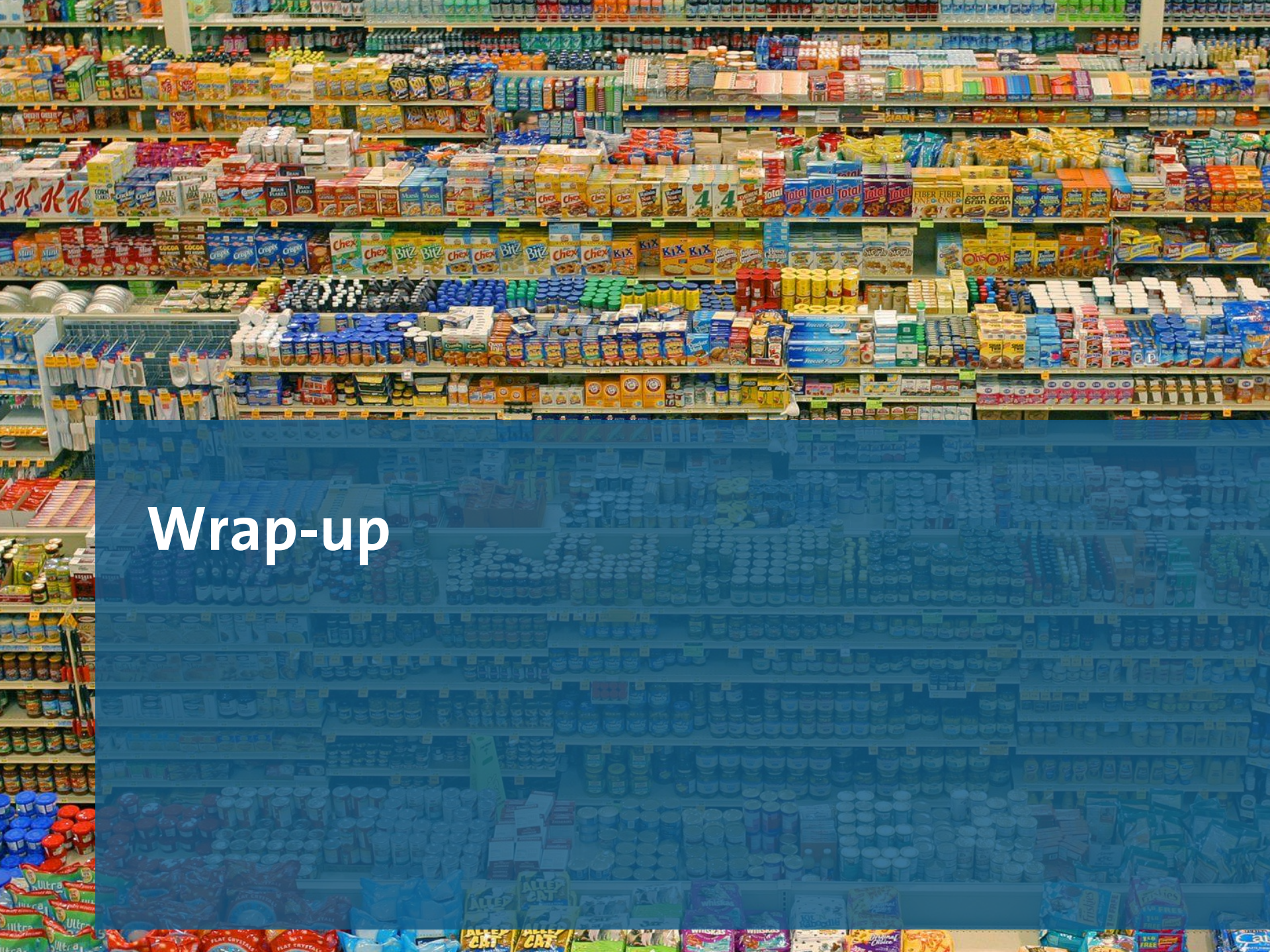
Topic 2: Federal LCA Commons Guidelines, Resources, and Tools

Guidelines	Tools	Resources
Data Submission Handbook	Collaboration Server	Electricity LCI
Elementary Flow Nomenclature	Unit Process Template	Elementary Flow Master List
LCI Data Quality Guidance		Federal Commons Data Portal
Technosphere Flow Nomenclature		LCI Data Quality Scheme

[1]Wesley W. Ingwersen, Ezra Kahn, Alberta Carpenter, Timothy Skone, Peter Arbuckle, Richard Bergman, Michael Wang, Joshua Kneifel, Kelly Scanlon, Federal LCA Commons Public Discussion Forum, presentation, LCA XVII, Portsmouth, 4 Oktober 2017.

LCA Collaboration Server, status 2

- currently ongoing: role-out, trainings, first hosted solutions
- we hope and expect the LCA Collaboration Server is an interesting offer to become a major element for any entity worldwide dealing with LCA data handling, be it
  - public agencies releasing and reviewing data
  - large teams jointly developing and using models (institutes, companies)



Wrap-up

GreenDelta

sustainability consulting + software

¡Muchas gracias!

Contact: Dr. Andreas Ciroth

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