

# Technical Helpdesk for National LCA Databases

---

## Training on Data Acquisition and Dataset Development – Parts 2 and 2a – Raw Data Acquisition

Content from: Andreas Ciroth, GreenDelta and Amir Safaei, ecoinvent

**Managed by SETAC**

---

### Disclaimer

The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the United Nations Environment Programme concerning the legal status of any country, territory, city or area or of its authorities, or concerning delimitation of its frontiers or boundaries. Moreover, the views expressed do not necessarily represent the decision or the stated policy of the United Nations Environment Programme, nor does citing of trade names or commercial processes constitute endorsement.

# Primary raw data acquisition

---

## Objective

- In this first practical exercise, we will create datasets from original sources, i.e. from “raw data”.
- Raw data = data that is not yet modeling in an inventory dataset with inputs and outputs matched to a reference flow\*

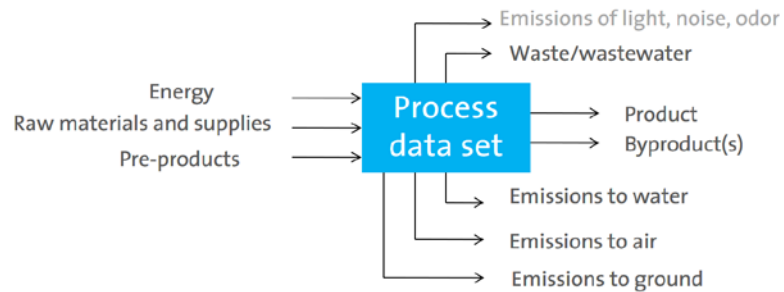
A formal definition of raw data is as follows:

- “Data used in unit process inventory modeling to deliver inventory data at the end, which are extracted from various data sources, such as bookkeeping of a plant, national statistics, or journal literature.”

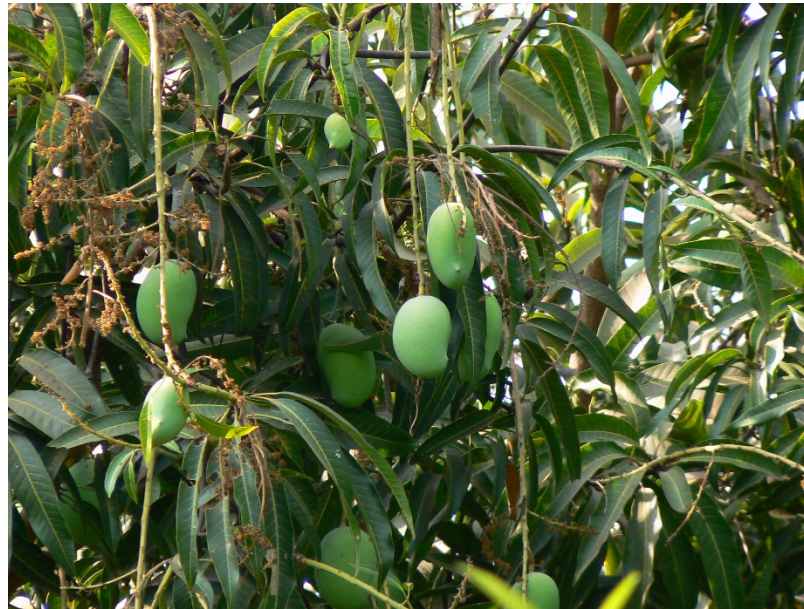
Source: Sonnemann, G. and B. Vigon, 2011. Global Guidance Principles for Life Cycle Assessment Databases, UN Environment Programme, ISBN: 978-92-807-3174-3, DTI/1410/PA

# Primary raw data acquisition

## Background



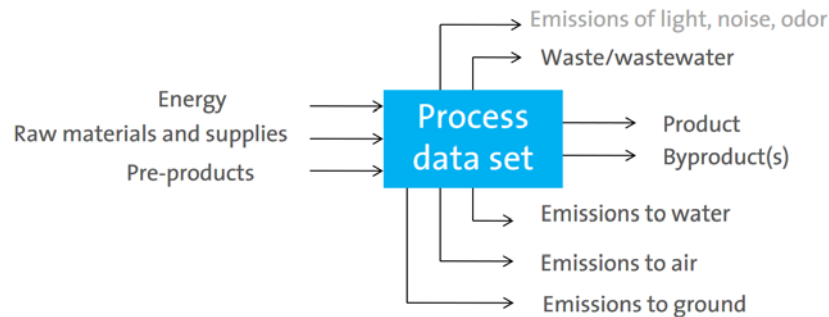
- Mango production with two processes



# Primary raw data acquisition

---

## Background



- Mango production with two processes
- Two groups:
  - 1) LCA modeler, data collectors
  - 2) Mango producer, data owners

# Primary raw data acquisition

---

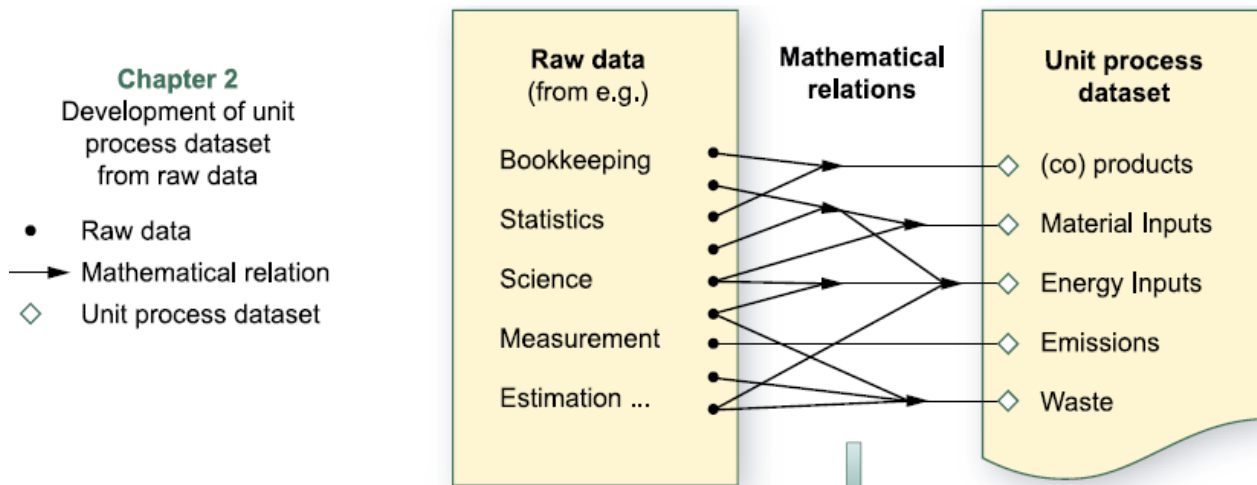
## Exercise 1-- Mango Production

- Producers - Own the data, exercise guidance filename (MSWord):  
**Mango\_data\_producer\_1<sup>st</sup>\_exercise\_ver8-17**
- Modelers - Collect the data, exercise guidance filename (MSWord):  
**Mango\_data\_collector\_modeler\_1<sup>st</sup>\_exercise\_ver8-17**
- Data collection template (generic), filename (MS Excel):  
**Data\_collection\_template\_generic\_ver8-17**

# Primary raw data acquisition

## Background

- Using “raw data” to create unit process datasets



Source: Wang, H., et al., Development of Unit Process Datasets, Chapter 2 in Sonnemann and Vigon (eds.) , 2011. Global Guidance Principles for Life Cycle Assessment Databases, UN Environment Programme, ISBN: 978-92-807-3174-3, DTI/ 1410/ PA.

# Primary raw data acquisition

## Background

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

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

Project

Intended applications

Data generator

Data set generator / modeller

Data entry by

Time stamp (last saved)

Data set format(s)

Converted original data set from

Data entry by

Official approval of data set by producer/operator

Publication and ownership

UUID

Data set version

Preceding Data set version

Permanent data set URI

Workflow and publication status

Unchanged re-publication of

Owner of data set

Copyright

Access and use restrictions

Inputs and Outputs

inputs

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	

# Primary raw data acquisition

## Background

- What does a process dataset in principle look like?

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

Process information

Modeling 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 by 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

Roosted

Data entry by

2012-03-28T17:40:25.397+02:00

Time stamp (last saved)

LCID format

Ecoprofiles

Converted original data set from

PE INTERNATIONAL

Data entry by

Plastics Europe

Official approval of data set by producer/operator

Publication and ownership

76d5aaa4-37e2-40b2-994c-03292b600074

USD

03 00 000

Preceding Data set version

Acrylonitrile-Butadiene-Styrene granulate (ABS): production mix, at plant

Permanent data set URL

<http://ica.jrc.ec.europa.eu/icaconfhub/datasets/elcd/processes/76d5aaa4-37e2-40b2-994c-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://ica.jrc.ec.europa.eu>. Please note e.g. that reference must be given 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

metadata

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	SAC0333.metal				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	HS0205000				-3.97796	-3.97796	0.0	0.0		-1.0 %	Mixed primary / secondary	Unknown derivation	

inputs

outputs



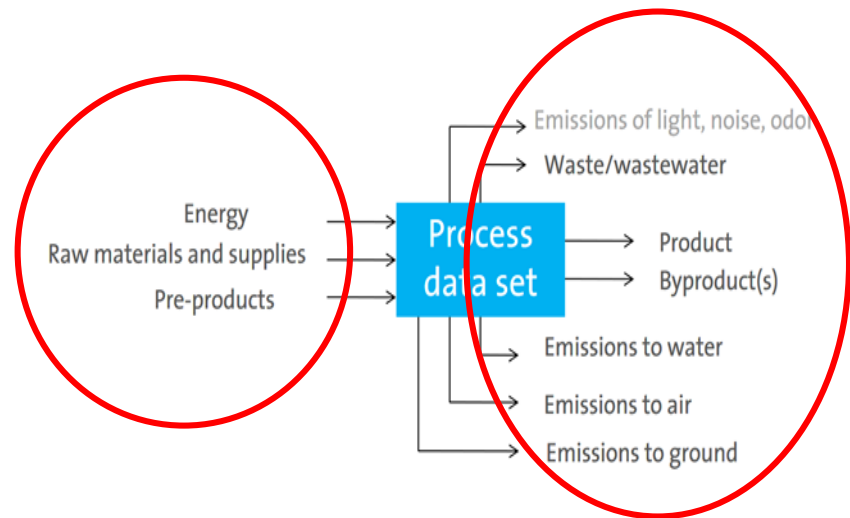
# Primary raw data acquisition

---

## The exercise

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

- ✓ Metadata?
- ✓ Inputs?
- ✓ Outputs?



# Primary raw data acquisition

---

## The exercise

### Training Workshop on Data Acquisition and Dataset Development for Life Cycle Inventory

#### Exercise 1 - Primary raw data acquisition and modeling

##### **Data for producers:<sup>1</sup>**

The dataset covers all relevant steps involved with a Mango production from cradle-to-gate, i.e. all processes from raw materials extraction till Mango harvesting are taken into account.

---

##### **PART 1.1 – Primary data collection**

You are a successful producer that started your 20ha Mango plantation about 7 years ago, for that you needed to start from scratch and buy a piece of land that still had a forest on it, thus, first of all it was necessary to clean the “fields”.

Preparing the field for the plantation demanded machines to cut the trees and prepare the fields, electricity to power some machinery, diesel and land.

The machines can clean and prepare 263 ha of field during its lifetime which means that for your Mango plantation only 0.076 Machine units were necessary.

# Primary raw data acquisition

---

## The exercise\*

Mango producer:

The information you have is in the exercise sheet.

You have of course an interest that your production appears environmental friendly.

\* This exercise is more effective if the roles of data owner/provider and collector/modeler are played separately.



# Primary raw data acquisition

---

## The exercise

**Data modeler:** The mango producer is your main data source for what is happening in this specific production, but you have of course a basic understanding of what is happening when growing Mangos.

**You want to model two processes:**

- *Field preparation*, output is the prepared area
- *Mango production on the field*, output is the Mango produced.

Simply create the processes in MS Excel, with inputs and outputs

# Primary raw data acquisition

## The exercise: example of an MS Excel sheet for data collection

In reality, two activities may be additionally needed in order to maximize the likelihood of getting complete and high quality data:

- Data collectors may need to establish a relationship with the data owners, possibly through a trade association, to help them understand the data needs better and to build credibility and
- The data collection template may need to be customized for the processes of interest. This minimizes the chances of data input errors and relieves data providers of the requirement to understand all the technicalities of the process.

Process		
Process identification <sup>(1)</sup>		
Process operator <sup>(2)</sup>		
Location <sup>(3)</sup>		
Quantitative reference and unit <sup>(4)</sup>		
Contact person <sup>(5)</sup>		
Address		
Telephone		
e-mail		
Process flowsheet <sup>(6)</sup>		
Inputs	Amount	Unit <sup>(9)</sup>
Energy source incl. efficiency <sup>(14)</sup>		
Material Inputs <sup>(15)</sup>		
Service Inputs <sup>(16)</sup>		
Outputs	Amount	Unit <sup>(9)</sup>
Product(s) <sup>(17)</sup>		
Emissions to air <sup>(18)</sup>		
Emissions to water <sup>(19)</sup>		

# Primary raw data acquisition

---

## The exercise - results

- The data collector groups are to present results
- The data owner groups are to describe their experience with supporting the collection effort\*

\* Oftentimes, data owners are given their own data (after modeling) to compare with the dataset average as a type of compensation for their cooperation.



# Primary raw data acquisition

---

## Elements in a data set, detail

- **General documentation considerations**
  - Name and classification – NACE or ISIC
  - Scope of the dataset
  - Functional unit and reference flow
  - Allocation information (if relevant)
  - Data quality information
  - Hints on interpretation
- **Specific requirements for documentation of unit process datasets**
  - Data source: raw data, mathematical relations, unit process dataset, supportive information
  - References and boundaries: description of goal and scope definition
  - Calculation models and other conventions:
  - Validation results

# Primary raw data acquisition

---

## Collecting information efficiently

- Data collection template\*
- (Data collection software, with template)

\*While it is always possible to use a generic data template, such as the supplied file for the trainee exercise, developing a process-specific template can have advantages. Such templates are more time-consuming to create, but once developed are more likely to make data providers more comfortable with the data request and are much less error prone. An example of such a customized template is:

[Data collection template\\_example\\_customized\\_May 2013](#)

Evaluate the pros and cons of a generic vs. custom template for your situation



# Primary raw data acquisition

---

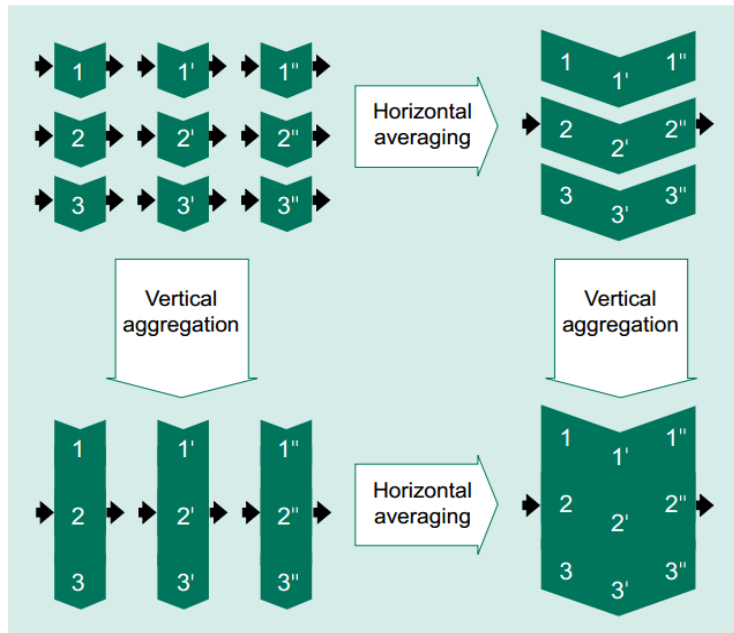
## Protecting sensitive information

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

# Primary raw data acquisition

## Protecting sensitive information

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



Source: Broadbent, C., et al.:  
Aggregated data development,  
Chapter 3 in Sonnemann and Vigon  
(eds.), 2011. Global Guidance Principles  
for LCA Databases, UN Environment  
Programme, ISBN: 978-92-807-3174-3,  
DTI/ 1410/ PA.

# Primary raw data acquisition

---

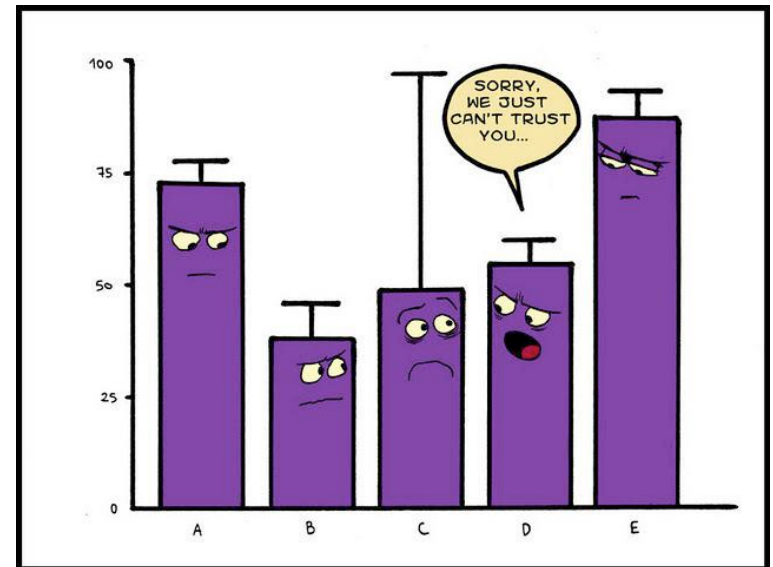
## Protecting sensitive information

- 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

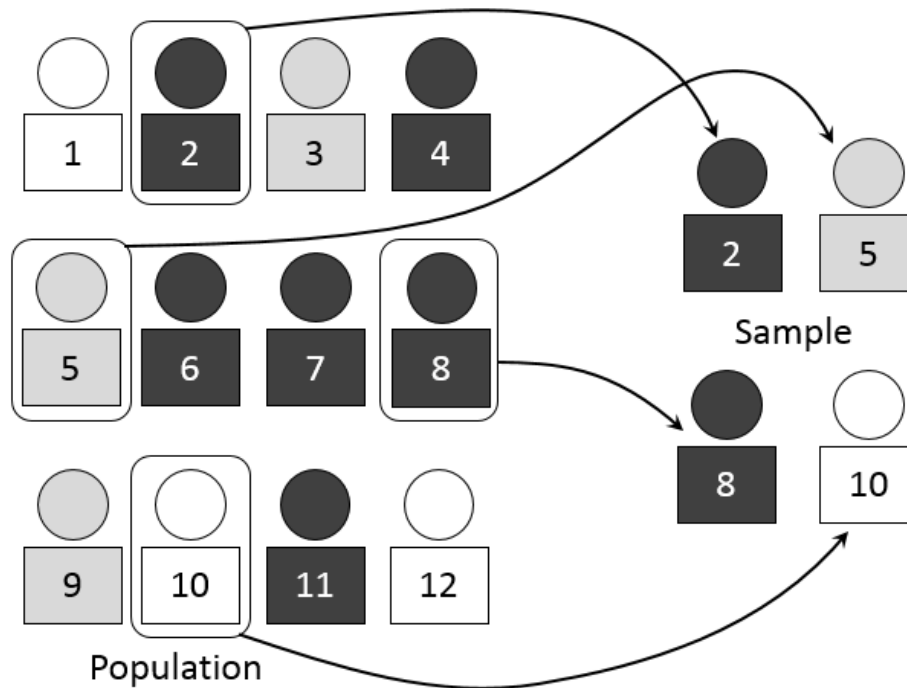
## How to create representative data sets

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



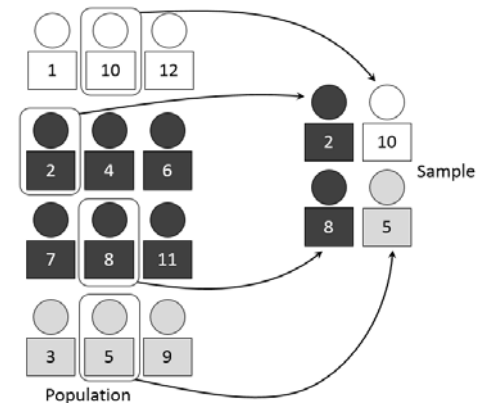
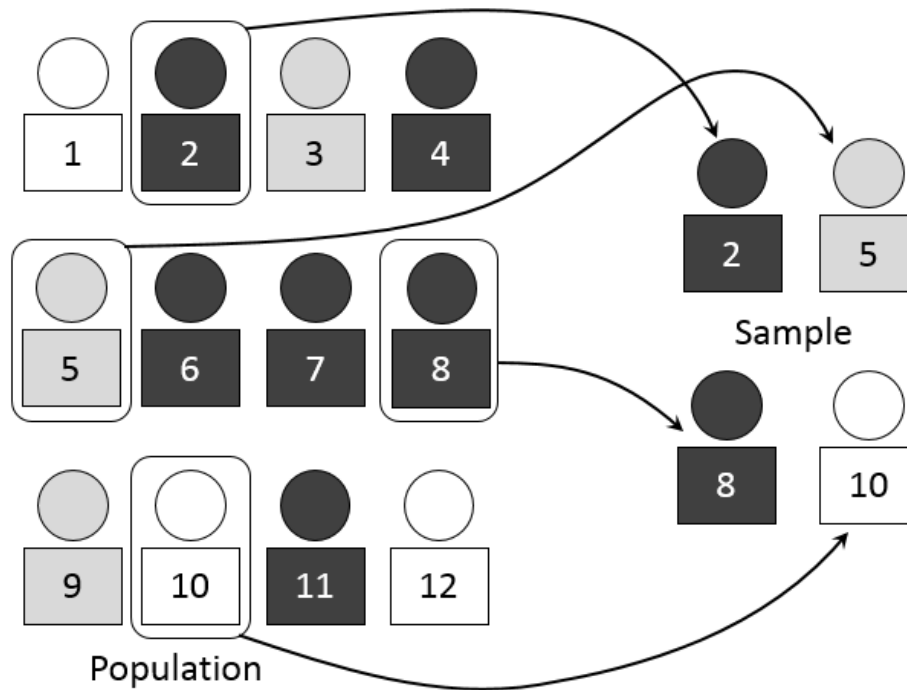
# Primary raw data acquisition

How to create representative data sets:  
statistical sampling



# Primary raw data acquisition

How to create representative data sets:  
statistical sampling

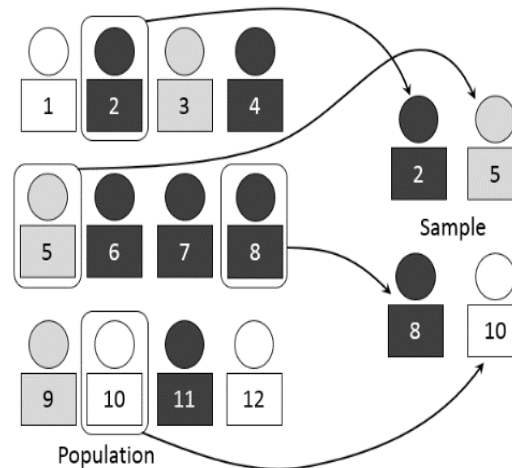
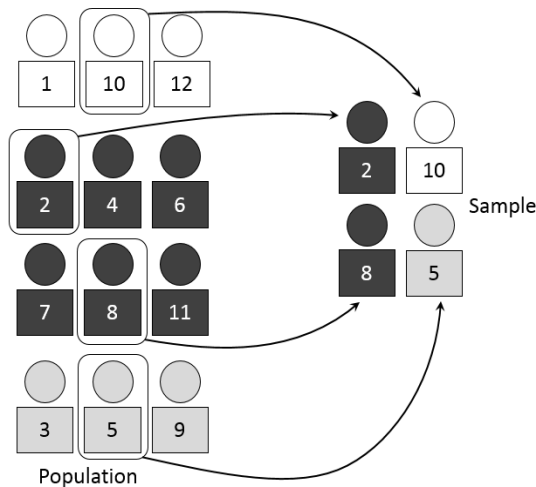


CC SA 4.0, Dan Kernler

# Primary raw data acquisition

## How to create representative data sets: statistical sampling

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



CC SA 4.0, Dan Kernler

# Technical Helpdesk for National LCA Databases

---

## Training on Data Acquisition and Dataset Development – Part 2a – Raw Data Acquisition, Documentation and Quality Assurance

Content from: Andreas Ciroth, GreenDelta and Amir Safaei, ecoinvent      **Managed by SETAC**

---

### **Disclaimer**

The designations employed and the presentation of the material in this document do not imply the expression of any opinion whatsoever on the part of the United Nations Environment Programme concerning the legal status of any country, territory, city or area or of its authorities, or concerning delimitation of its frontiers or boundaries. Moreover, the views expressed do not necessarily represent the decision or the stated policy of the United Nations Environment Programme, nor does citing of trade names or commercial processes constitute endorsement.



# Primary raw data acquisition, 2a

---

Now we have collected data..

- What's next:
  - Input of data into a common form
  - Quality assurance (QA)
  - Documentation

**QUALITY** 

# Primary raw data acquisition, 2a

---

## Data input and compilation

- MSExcel (generic or custom)
- Specific solutions
  - Ecoeditor: [www.ecoinvent.org/data-provider/data-provider-toolkit/eceditor/eceditor.html](http://www.ecoinvent.org/data-provider/data-provider-toolkit/eceditor/eceditor.html)
  - limesurvey query tool, configured (open source survey tool, web-based)
  - Excel template for openLCA
  - ...
- LCA software:
  - ..

# Primary raw data acquisition, 2a

---

## Data input and compilation

- LCA software:
  - openLCA
  - openLCA, with Collaboration Server ([www.openlca.org/collaboration-server/](http://www.openlca.org/collaboration-server/))
  - openLCA, with soda4LCA hub
  - SimaPro
  - SimaPro, with share&collect
  - GaBi
  - GaBi, with LCA hub
  - ...

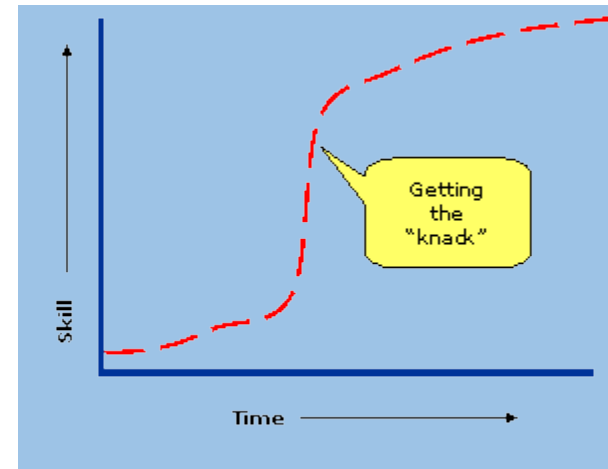


# Primary raw data acquisition, 2a

---

## Data input and compilation

- Considerations in selecting :
  - ✓ Consistency
  - ✓ Costs and effort
  - ✓ Ease of use
  - ✓ Learning curves



# Primary raw data acquisition, 2a

---

## Data verification and quality assurance

- Did you collect or did you receive “correct” datasets?

# Primary raw data acquisition, 2a

---

## Data verification and quality assurance

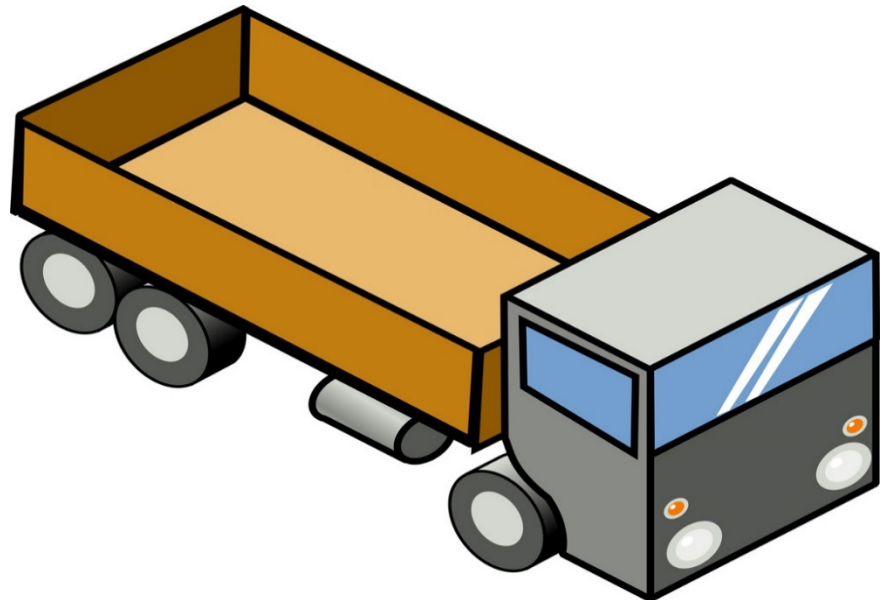
- Did you collect, did you receive “correct data sets”?
  - What would you expect?
    - Missing flows
    - Unnecessary flows
    - Units plausible
  - Mass balance, energy balance
  - Relations between flows
    - water in, water out
    - SO<sub>2</sub> output by fuel S-content

# Primary raw data acquisition, 2a

---

QA, real processes (already in databases)  
Light commercial truck transport, USA

- What would you expect?



# QA, real processes

From USLCA  
Commons  
database,  
2004

P Transport, light commercial truck, diesel powered, Northeast - RNA ☒

Process: Transport, light commercial truck, diesel powered, Northeast

## Inputs

Flow	Category	Amount	Unit
F Diesel, at refinery - RNA	Product flows	0.276571322110192	L
F Transport, barge, average fuel mix - ...	Product flows	0.0862336594325369	t*km
F Transport, combination truck, avera...	Product flows	0.089154960258969	t*km
F Transport, ocean freighter, average ...	Product flows	0.609499460075728	t*km
F Transport, pipeline, unspecified petr...	Product flows	0.8986512409252	t*km
F Transport, train, diesel powered - R...	Product flows	0.166374128034546	t*km

## Outputs

Flow	Category	Amount	Unit
F Ammonia	Emission to air/unspecified	2.39879420774344E-5	kg
F Carbon dioxide, fossil	Emission to air/unspecified	0.72614143603924	kg
F Carbon monoxide, fossil	Emission to air/unspecified	0.00312361727830329	kg
F Dinitrogen monoxide	Emission to air/unspecified	3.01784450574847E-6	kg
F Hydrocarbons (other than methane)	Emission to air/unspecified	7.02201044264229E-4	kg
F Methane	Emission to air/unspecified	2.02720478788059E-5	kg
F Nitrogen dioxide	Emission to air/unspecified	3.77984670306497E-4	kg
F Nitrogen oxide	Emission to air/unspecified	0.00427380226811679	kg
F Nitrogen oxides	Emission to air/unspecified	0.00465179641221559	kg
F Particulates, < 10 um	Emission to air/unspecified	2.22657367961114E-5	kg
F Particulates, < 10 um	Emission to air/unspecified	2.84157349924053E-4	kg
F Particulates, < 10 um	Emission to air/unspecified	7.68309810944433E-6	kg
F Particulates, < 2.5 um	Emission to air/unspecified	2.75643935778848E-4	kg
F Particulates, < 2.5 um	Emission to air/unspecified	5.82872478219123E-6	kg
F Particulates, < 2.5 um	Emission to air/unspecified	1.84246990550094E-6	kg
F Sulfur dioxide	Emission to air/unspecified	1.23298322287085E-5	kg
F VOC, volatile organic compounds	Emission to air/unspecified	7.25107114851339E-4	kg
F Transport, light commercial truck,...	Product flows	1.0	t*km



# QA, real processes

1. 1 Diesel / 100km?
2. S content in Diesel?
3. CO2 emissions in relation to Diesel consumption?
4. Mass balance?
5. Inputs?
6. Other aspects?

P Transport, light commercial truck, diesel powered, Northeast - RNA

Process: Transport, light commercial truck, diesel powered, Northeast

## Inputs

Flow	Category	Amount	Unit
F Diesel, at refinery - RNA	Product flows	0.276571322110192	L
F Transport, barge, average fuel mix - ...	Product flows	0.0862336594325369	t*km
F Transport, combination truck, avera...	Product flows	0.089154960258969	t*km
F Transport, ocean freighter, average ...	Product flows	0.609499460075728	t*km
F Transport, pipeline, unspecified petr...	Product flows	0.8986512409252	t*km
F Transport, train, diesel powered - R...	Product flows	0.166374128034546	t*km

## Outputs

Flow	Category	Amount	Unit
F Ammonia	Emission to air/unspecified	2.39879420774344E-5	kg
F Carbon dioxide, fossil	Emission to air/unspecified	0.72614143603924	kg
F Carbon monoxide, fossil	Emission to air/unspecified	0.00312361727830329	kg
F Dinitrogen monoxide	Emission to air/unspecified	3.01784450574847E-6	kg
F Hydrocarbons (other than methane)	Emission to air/unspecified	7.02201044264229E-4	kg
F Methane	Emission to air/unspecified	2.02720478788059E-5	kg
F Nitrogen dioxide	Emission to air/unspecified	3.77984670306497E-4	kg
F Nitrogen oxide	Emission to air/unspecified	0.00427380226811679	kg
F Nitrogen oxides	Emission to air/unspecified	0.00465179641221559	kg
F Particulates, < 10 um	Emission to air/unspecified	2.22657367961114E-5	kg
F Particulates, < 10 um	Emission to air/unspecified	2.84157349924053E-4	kg
F Particulates, < 10 um	Emission to air/unspecified	7.68309810944433E-6	kg
F Particulates, < 2.5 um	Emission to air/unspecified	2.75643935778848E-4	kg
F Particulates, < 2.5 um	Emission to air/unspecified	5.82872478219123E-6	kg
F Particulates, < 2.5 um	Emission to air/unspecified	1.84246990550094E-6	kg
F Sulfur dioxide	Emission to air/unspecified	1.23298322287085E-5	kg
F VOC, volatile organic compounds	Emission to air/unspecified	7.25107114851339E-4	kg
F Transport, light commercial truck,...	Product flows	1.0	t*km

# Primary raw data acquisition, 2a

---

## Data documentation

- Author
- Date
- Aim of the data set
- Sources
- Potential modifications of the data set
- Limitations
- ...

# Primary raw data acquisition, 2a

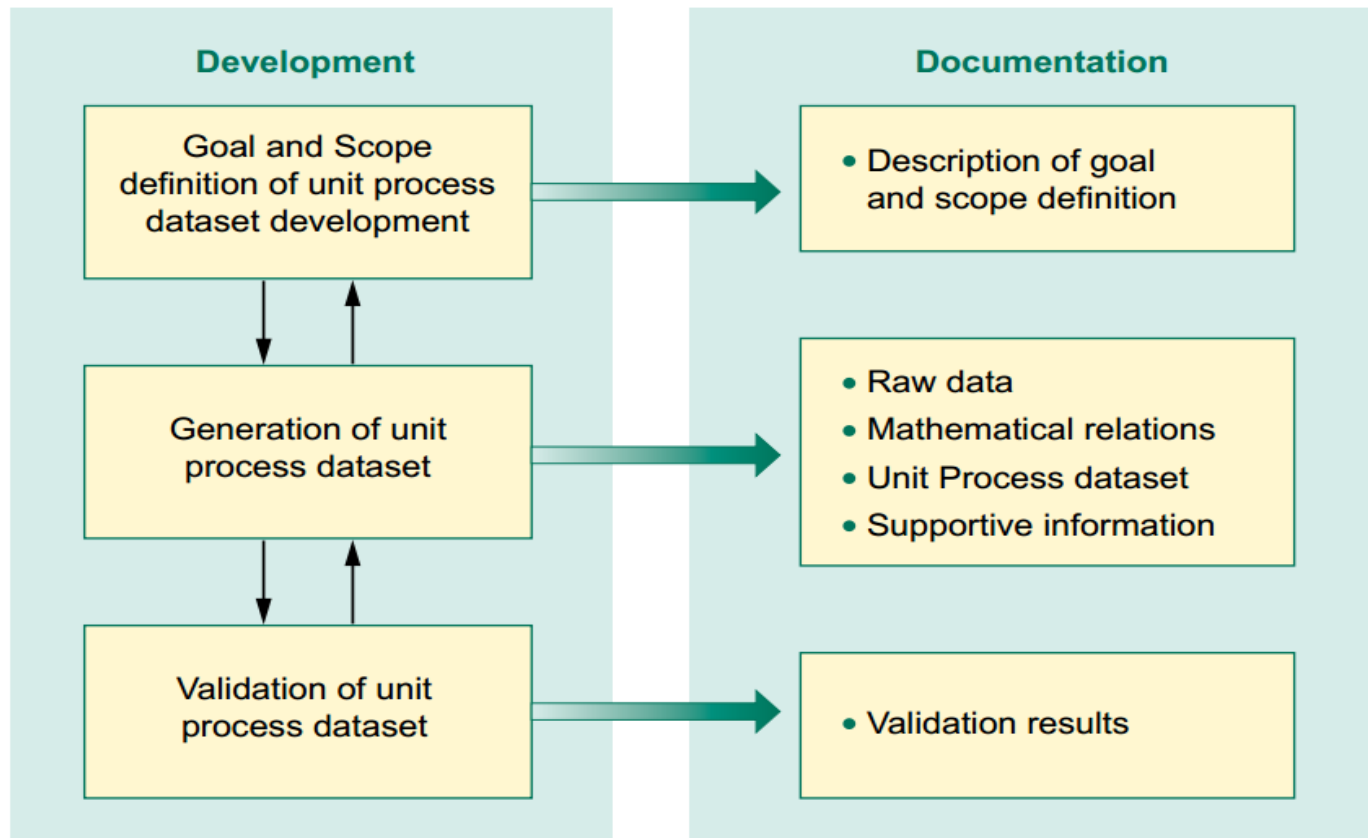
---

## Data documentation

- Author
  - Date
  - Aim of the data set
  - Sources
  - Potential modifications of the data set
  - Limitations
  - ...
- 
- At which level?
    - Dataset
    - Individual input/output?

# Primary raw data acquisition, 2a

## Data documentation



Source: Wang, H., et al., Development of Unit Process Datasets, Chapter 2 in Sonnemann and Vigon (eds.), 2011. Global Guidance Principles for Life Cycle Assessment Databases, UN Environment Programme, ISBN: 978-92-807-3174-3, DTI/ 1410/PA.

# Primary raw data acquisition, 2a

---

Finally, “metaindicators”

Metaindicators = metadata descriptors

→ everything that describes a process to understand “fitness for purpose” (data quality) for a process data set.



# Metaindicators

UNEGGLAD WG 3  
metadata  
descriptors, task 3,  
final draft report,  
June 2017

See also Part 2b  
on Meta-  
indicators and  
GLAD

		Goal	Value & representation	Conformance
ID	Process name		0a	
	Process type		0b	
gvc Descriptors	Time	Ia	IVa	IIa
	Geography	Ib	IVb	IIb
	Technology	Ic	IVc	IIc
	Model completeness	Id	IVf	IId
	Sample representativeness	Ie	IVg	IIe
	LCA nomenclature systems		IVd	
	LCIA methods	Ig	IVe	
Modeling	LCI modeling type		IVh	
	System boundaries		IVi	
	Multifunctional processes		IVj	
	Biogenic carbon		IVk	
	Land use		IVl	
	Wastes and end-of-life		IVm	
	Water		IVn	
	Infrastructure/capital goods		IVo	
	Long-term emissions		IVp	
	Temporary carbon storage		IVq	
Sampling	Sample approach	If		
	Reliability of the sources used		IIla	
Calculation	Aggregation type if any		Vla	
	Data set review performed		Va	
	Type of data set review		Vb	
	Quality assurance performed		Vc	
	Reviewing person(s)		Vd	
QA				
Administrative	Copyright protected data set?		VIIa	
	Copyright holder		VIIb	
	Free data set or for purchase?		VIIc	
	Data set license		VIIId	
	Data set contact		VIIe	

Descriptor element supported in / provided by

GLAD	ILCD	ILCD & EcoSpold02	EcoSpold02	not applicable	not foreseen	(ILCD)	(ILCD & EcoSpold02)	(EcoSpold02)
------	------	-------------------	------------	----------------	--------------	--------	---------------------	--------------

# Primary raw data acquisition, 2a

---

Exercise, continued:

Do you need to add any metadata to your process, as an LCA modeler?

As producer, can you tell if additional metadata that are needed?

# For helpdesk assistance –

---

- Become a Helpdesk member:
  - To access the Helpdesk exchange space (or any other Clearinghouse area), you will need to create an account in the Clearinghouse ([www.scpclearinghouse.org](http://www.scpclearinghouse.org)):
  - Toward the bottom of the homepage you will see a button labeled ‘Join the Community now’. Click on this link and open a form to allow you to create a login and profile.
  - Once logged in, you can modify or update your profile or explore the various SCP topic areas.
  - Go to ‘About’ and then to ‘Exchange Spaces’ where you will see Lifecycle Approaches in the drop down menu and one menu level below the Technical Helpdesk.
  - The Technical Helpdesk space will be available to any visitor, logged in or not. Without being logged in and joining the helpdesk space, any visitor can look at the various sections of the helpdesk space, but cannot contribute any content.
  - In order to become a member of the helpdesk space, on the homepage under the summary, is “Request space membership”. Click here, you will automatically be given rights of a members to contribute content, since it is a public group.
  - For your next login, you go directly to <http://spaces.scpclearinghouse.org/> and then choose the Technical Helpdesk space in the dropdown list.
- Helpdesk Manager - Bruce Vigon, Consultant to SETAC,
- Helpdesk Coordinator – Kristina Bowers, UN Environment, Economy Division