Task 3: Core meta-data descriptors and guidance on populating descriptors

Global LCA Access

WG3 of the Global Life Cycle Data Access Network (GLAD)

Andreas Ciroth, GreenDelta GmbH, ciroth@greendelta.com
Peter Arbuckle, Edivan Cherubini, Cassia Ugaya, Ashley Edelen

Release version 1.0 (internal version 11.3)
June 2017

Statement of Objectives[1]

The "Global LCA Data Access" network (GLAD) aims towards a global network comprised of independently operated and interoperable LCA databases, connecting multiple data sources to support life cycle assessment (LCA) in a way that facilitates sustainability-related decisions. GLAD will enable access by users to various LCA databases and their seamless delivery into LCA software, with sufficient documentation of metadata that allows for assessment of "fitness for purpose" for any end user. One of the main aspects of GLAD is the definition of metadata descriptors. The metadata descriptors are the basis for searching, filtering and sorting across registered datasets in GLAD, and a common understanding in the way they are documented will facilitate conversion of datasets between exiting life cycle inventory formats. As such, a common set of metadata descriptors will facilitate interoperability between datasets and allow to determine fitness for purpose of datasets by users. A working group set up under GLAD has identified a list of key metadata descriptors and definitions with the aim to make them applicable through existing and future LCA databases. The proposed descriptors need to fit to GLAD's ambition to deliver by 2017 an electronic system and protocol enabling access by users to the majority of the LCA databases; therefore a key criterion for this common set of descriptors is that they should be already contained in existing databases. It should be noted that even when the proposed descriptors are all contained in databases, their interpretation and documentation is not always consistent across databases; when this happens the burden on the final user to compare and check such metadata descriptors will be higher. GLAD will continue to work with database providers and related stakeholders to improve the consistency of all these descriptors in the future, following the guidance initiated in the second part of this document.

Structure of this report

The first part of this report provides a minimum set of metadata descriptors, which are readily applicable by the main existing LCA databases. These descriptors constitute the basic metadata descriptors required from datasets to be connected to GLAD in order to allow for interoperability.

The second part of this report sets out approaches for improving interoperability of datasets in the future. With a view to improving assessment of fitness for purpose, the report provides guidance on what kind of information should ideally be included in the metadata descriptors to allow for greater interoperability and better assessment of fitness for purpose.

Acknowledgements

UN Environment would like to thank the following experts who provided feedback on the report; their review does not imply endorsing the content of the report itself.

In alphabetical order:

Martin Baitz (Germany), Ashley Edelen (USA), Cristian Emhart (Chile), Rolf Frischknecht (Switzerland), Wesley Ingwersen (USA), Brandon Kuczenski (USA), Nongnuch Poolsawad (Thailand), Cristobal Loyola (Chile), Koichi Shobatake (Japan), Diogo A. Lopes Silva (Brazil), Kiyotaka Tahara (Japan), Bo Weidema (Denmark), Gregor Wernet (Switzerland), Marc-Andree Wolf (Germany)

[1] This first page with statement of objectives and acknowledgements was added in November 2017 for the on-line publication of the report. The rest of the report was approved in June 2017.

The descriptor short list

Per suggestion of the UN Environment Life Cycle Initiative secretariat, the short list of descriptors that is developed in the course of this report, is presented here in advance of the text. The list contains those fitness for purpose descriptor elements that were found, in a test of two real data sets in EcoSpold2 and ILCD, to be supported by these formats, i.e. it was possible to provide information for these fields. It was not considered for the evaluation whether and how far the provided information found in the data sets is contributes to an understanding of fitness for purpose.

When using this list, please be aware that the list was obtained by a mere technical procedure, finding common data format fields and related information in the two investigated datasets. It therefore cannot claim to summarise the information that is important for understanding fitness for purpose for datasets, nor can it claim to contain all the information necessary for making a dataset interoperable.

Short list of proposed metadata descriptors that should be available in process datasets to join the GLAD platform

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
0	Dataset identification						
a	Process name	General descriptive name of the process and/or its main good(s) or service(s) and/or its level of processing.	Dataset	Ad	DaP	Value	Example: corn grain; average tillage practice mix; at farm; 15% moisture The dataset name ideally conforms to a harmonized nomenclature convention which could be specified in the GLAD nomenclature group. An example could be: Base name; treatment, standards, routes; mix and or location type; quantitative product or process properties
b	Process type	Unit process or aggregated process	Dataset	Ad	DaP	Value	
I	Goal and scope						
d	Reference model completeness	Targeted coverage of inputs and outputs in the dataset	Dataset	FI	DaP	Goal	In terms of numbers of flows and /or a complete coverage of the amounts, leading to a complete mass and / or energy balance
e	Reference sample representativeness	Targeted sample representativeness, in line with the foreseen sampling approach	Dataset	FI	DaP	Goal	The intended sampling approach is If)
f	Intended sample approach	Scientific or expert-based sampling	Dataset	FI	DaP	Goal	Or: expert judgement as default
<u>IV</u>	<u>Life cycle and model</u>						
a	Time	Time or time period relevant	Exchange	FI	DaP	Value	
b	Geography	Geographical area or point relevant	Exchange	FI	DaP	Value	
c	Technology	Technology or technology mix relevant	Exchange	FI	DaP	Value	

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
d	Supported LCA nomenclature system(s)	See name	Dataset	Fl	DaP	Value	ILCD reference flow list, ILCD 1.1 from May 2015; Ecoinvent 3.3 master data
g	Representativeness	For science-based sampling, variation coefficient plus documentation; for expert judgement, representativeness classes estimates	Exchange	St	DaP	Value	Static descriptor since it seems always desirable to have a representative dataset
h	LCI modeling approach	Attributional or consequential modeling	Dataset	Fl	DaP	Value	See also detailed format description, chapter 4
j	Method used to deal with multifunctional processes	Method used to assign the environmental burdens to the joint production of the reference flows.	Dataset	Fl	DaP	Value	
k	Biogenic carbon	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).	Dataset	Fl	DaP	Value	
V	Verification and quality assurance						
a	Dataset review performed	See name	Dataset	Fl	DaP	Value	
d	Reviewing person(s)	Person who conducted the review	Dataset	Fl	DaP	Value	
VI	Calculation						
a	Aggregation type if any	For an aggregated dataset, specify how the aggregation was performed.	Dataset	Fl	DaP	Value	Horizontal, along the supply chain / vertical, across several processes delivering similar products, mixed; partial or complete – so, 6 cases)
VII	VII Administrative						
a	Copyright protected dataset?	See name	Dataset	Ad	DaP	Value	
b	Copyright holder	See name	Dataset	Ad	DaP	Value	

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
c	Free dataset or for purchase?	See name	Dataset	Ad	DaP	Value	
d	Dataset license	See name	Dataset	Ad	DaP	Value	
e	Dataset contact	See name	Dataset	Ad	DaP	Value	

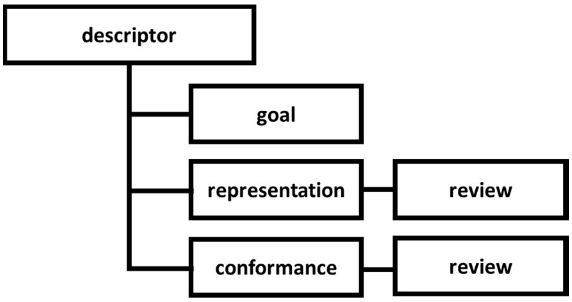
*Ad: administrative; St: static; Fl: flexible
**DaP: data provider

0. Executive summary

In summary, this report contains the following results:

Metadata descriptor structure

In the context of LCA data, each descriptor has several elements: goal, representation, and conformance; representation and conformance can be reviewed.



In this structure, **goal** means a statement about the goal for the descriptor, for the respective dataset. For example, for the topic time, a possible goal can be: Goal for this dataset is to represent the situation for the year 2015. **Representation** means information about what the dataset represents, for the specific descriptor. Again for time, a possible information for representation could be: This dataset is representing inputs and outputs valid for 2013. Finally and third, a **conformance** assessment shows how far the dataset deviates from the stated goal, for the descriptor. Taking again time as example, a possible conformance assessment result could be: The dataset is 2 years older than foreseen.

Although implicitly addressed in various modern LCA formats, it is the first time that these descriptor elements are mentioned and used in the LCA context. The "goal, representation, conformance" concept provides a critical foundation for data set provision and use in GLAD.

The GLAD initiative is a new model for providing access to globally distributed datasets (as opposed to centrally curated), which also represent a diversity of goals, scopes, and system boundary conditions. GLAD's quality and success is predicated upon a data provider's ability to fully document the data set and the user's ability to fully understand its appropriate application. The "goal, representation, conformance" approach provides the structure, to enable data providers and users to share datasets through GLAD.

Proposal for metadata descriptors distinguished in goal, representation and conformance

A list of metadata descriptors is proposed in the text, it is distinguished in goal, value and representation, and conformance, for each descriptor.

Guidance on populating the descriptors. The text recommends rules for "populating" the descriptors, and thereby also addresses what should not be entered. The recommendations are:

1. Using *flat descriptors*: Cross-references, from principles to deviations from principles to further sources, make it very difficult to understand the actual representation of a descriptor for a dataset. For the descriptors, a flat result must be available, which condenses the possibly various sources of input for a descriptor into the actual representation for the descriptor.
2. *Avoidance of redundant and superfluous information*. Descriptors should not contain more information than necessary; specifically, only the sources used in the dataset should be listed, and it should be clear which information has been taken from which source.
3. The *descriptor structure* should follow the structure of goal and scope / representation / conformance, and for each of these descriptor segments, this should be clear for the user.

4. *Evaluation statements* should only be placed in the conformance segment of a descriptor to avoid misleading claims. Descriptions should be separated from the evaluation, and be in the representation segment. Terms like “relevant”, “quality”, “consistent” are evaluation statements.

Further, aspects for the data format of the descriptors are proposed. A detailed format specification is not provided, though, because descriptors often are already contained in data formats and software. A detailed specification is seen as relevant in the implementation.

The proposed descriptors elements are to be provided in part by the data provider, and in part by GLAD users and by the GLAD system. For time, geography, and technology, the assessment of fitness for purpose is fully case dependent; therefore, for these descriptors, users enter the respective target representation, and the dataset conformance is assessed using the values for time, geography, and technology stored in the database. This assessment is to be performed in the GLAD system, and its result, the conformance for, “delivered” to the user by the GLAD system0.

The proposed metadata descriptors have been **tested on real datasets** in EcoSpold02 and ILCD format; in general, both formats support the proposed descriptors well, with few exceptions.

		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		IIIa	
Calculation	Aggregation type if any		VIa	
QA	Data set review performed		Va	
	Type of data set review		Vb	
	Quality assurance performed		Vc	
	Reviewing person(s)		Vd	
Administrative	Copyright protected data set?		VIIa	
	Copyright holder		VIIb	
	Free data set or for purchase?		VIIc	
	Data set license		VIId	
	Data set contact		VIIe	

Descriptor element supported in / provided by

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

(Format in bracket): partial support:

Based on the analysis of the availability of the proposed descriptors, a shorter list of descriptors can be proposed, which is summarized in Figure 0. In this figure, as in the figures above, a green background indicates data that should not be stored at the dataset but instead entered by a GLAD user and in turn calculated by the GLAD system.

	Goal	Value & representation	Conformance
Process name		0a	
Process type		0b	
Time	Ia	IVa	IIa
Geography	Ib	IVb	IIb
Technology	Ic	IVc	IIc
Model completeness	Id		
Sample representativeness	Ie	IVg	
LCI modeling type		IVh	
Multifunctional processes		IVj	
Biogenic carbon		IVk	
Sample approach	If		
Aggregation type if any		VIa	
Data set review performed		Va	
Reviewing person(s)		Vd	
Copyright protected data set?		VIIa	
Copyright holder		VIIb	
Free data set or for purchase?		VIIc	
Data set license		VIId	
Data set contact		VIIe	

Figure 0: Short list of descriptors

The short list of descriptors is further detailed in Table 0 in the next page, together with their description. Table 0 is an extract of Table 1 and includes only those metadata descriptors that are part of the short list.

Table 0: Short list of proposed metadata descriptors that should be available in process datasets to join the GLAD platform

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
0	Dataset identification						
a	Process name	General descriptive name of the process and/or its main good(s) or service(s) and/or its level of processing.	Dataset	Ad	DaP	Value	Example: corn grain; average tillage practice mix; at farm; 15% moisture The dataset name ideally conforms to a harmonized nomenclature convention which could be specified in the GLAD nomenclature group. An example could be: Base name; treatment, standards, routes; mix and or location type; quantitative product or process properties
b	Process type	Unit process or aggregated process	Dataset	Ad	DaP	Value	
I	Goal and scope						
d	Reference model completeness	Targeted coverage of inputs and outputs in the dataset	Dataset	FI	DaP	Goal	In terms of numbers of flows and /or a complete coverage of the amounts, leading to a complete mass and / or energy balance
e	Reference sample representativeness	Targeted sample representativeness, in line with the foreseen sampling approach	Dataset	FI	DaP	Goal	The intended sampling approach is If)
f	Intended sample approach	Scientific or expert-based sampling	Dataset	FI	DaP	Goal	Or: expert judgement as default
IV	Life cycle and model						
a	Time	Time or time period relevant	Exchange	FI	DaP	Value	
b	Geography	Geographical area or point relevant	Exchange	FI	DaP	Value	
c	Technology	Technology or technology mix relevant	Exchange	FI	DaP	Value	
d	Supported LCA	See name	Dataset	FI	DaP	Value	ILCD reference flow list, ILCD 1.1 from May

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
	nomenclature system(s)						2015; Ecoinvent 3.3 master data
g	Representativeness	For science-based sampling, variation coefficient plus documentation; for expert judgement, representativeness classes estimates	Exchange	St	DaP	Value	Static descriptor since it seems always desirable to have a representative dataset
h	LCI modeling approach	Attributional or consequential modeling	Dataset	FI	DaP	Value	See also detailed format description, chapter 4
j	Method used to deal with multifunctional processes	Method used to assign the environmental burdens to the joint production of the reference flows.	Dataset	FI	DaP	Value	
k	Biogenic carbon	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).	Dataset	FI	DaP	Value	
V	Verification and quality assurance						
a	Dataset review performed	See name	Dataset	FI	DaP	Value	
d	Reviewing person(s)	Person who conducted the review	Dataset	FI	DaP	Value	
VI	Calculation						
a	Aggregation type if any	For an aggregated dataset, specify how the aggregation was performed.	Dataset	FI	DaP	Value	Horizontal, along the supply chain / vertical, across several processes delivering similar products, mixed; partial or complete – so, 6 cases)
VII	VII Administrative						
a	Copyright protected dataset?	See name	Dataset	Ad	DaP	Value	
b	Copyright holder	See name	Dataset	Ad	DaP	Value	
c	Free dataset or for	See name	Dataset	Ad	DaP	Value	

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Comment / Example
	purchase?						
d	Dataset license	See name	Dataset	Ad	DaP	Value	
e	Dataset contact	See name	Dataset	Ad	DaP	Value	

*ad: administrative; st: static; fl: flexible
**DaP: data provider

Content

List of figures

List of tables

1. Motivation and starting points

This is the second report of the Working Group (WG) Metadata Descriptors; the first, called "Meta-Data Needs Assessment – Element 1" investigated the need for assessing fitness for purpose for LCA datasets.

The aim of this text is to:

- Identify key metadata descriptors which determine fitness for purpose for Life Cycle Assessment (LCA) datasets;
- Characterise and define these key descriptors, with the aim to make them applicable through various data systems
- Provide guidance and assistance in completing and "populating" the identified metadata descriptors.

The following documentation was used to select and define descriptors:

- Working groups and round tables
  - Glad network reports (Glad WG1 2016)
  - The 1992 SETAC workshop in Leiden (SETAC 1992)
  - 1994 SETAC Data Quality Workshop in Pensacola (SETAC 1994)
- Standards
  - ISO standards and Technical Specifications (e.g., 14040, 14044, ISO 14048/TS)[2]
  - The Shonan Guidance Principles text of UNEP and SETAC (UNEP 2011)
- Recent work on LCA dataset review (Ciroth et al. 2015, Ciroth et al. 2016)
- Current practices, including common use of descriptors in current schemas, databases and formats
  - Data formats (e.g., ILCD; EcoSpold01, EcoSpold02, 2LD[3])
  - Databases (e.g., ELCD; ecoinvent; GaBi)
  - LCA data quality schemes (e.g., ILCD, ecoinvent, GaBi, US EPA)

Metadata was defined by the Global Guidance Principles for LCA Databases (Shonan Guidance Principles) in 2011 and is defined by ISO/IEC 1179-1:2004 as,

"data that defines and describes other data and processes."

For the purpose of clarity, this report further clarifies LCA metadata as and part of an LCA dataset which is not the input and output list of exchanges, their amounts and units. Furthermore, following the idea of modern dataset formats such as ILCD, this definition of metadata includes the process dataset (e.g. process name) and flow datasets.

1.1. WG 1 in GLAD, architecture and network

The Architecture and Network WG 1 in GLAD developed a "staircase" figure to visualise different levels of interoperability for LCA datasets, the highest being 15, "scientific quality level". This figure was broadly discussed and it is included in the GLAD Statement of Objectives. It was generally accepted that GLAD datasets should at least meet level 10, which is specified as

[2] ISO 14040 and 14044 do not deal with datasets directly, but they address many topics which are relevant for LCA datasets metadata (geography, time, uncertainty, data quality for example)

[3] 2LD is the LCA linked data format, developed by Michael Srocka of GreenDelta, as the core format of the openLCA software. It was initially developed in a project with US EPA and is now employed e.g. in the collaboration server which is in use at USDA and the LCACommons

> "dataset fields sufficiently correct with regards to interoperable unit process / product system, name, categorization and meta data" (Figure 1, below).

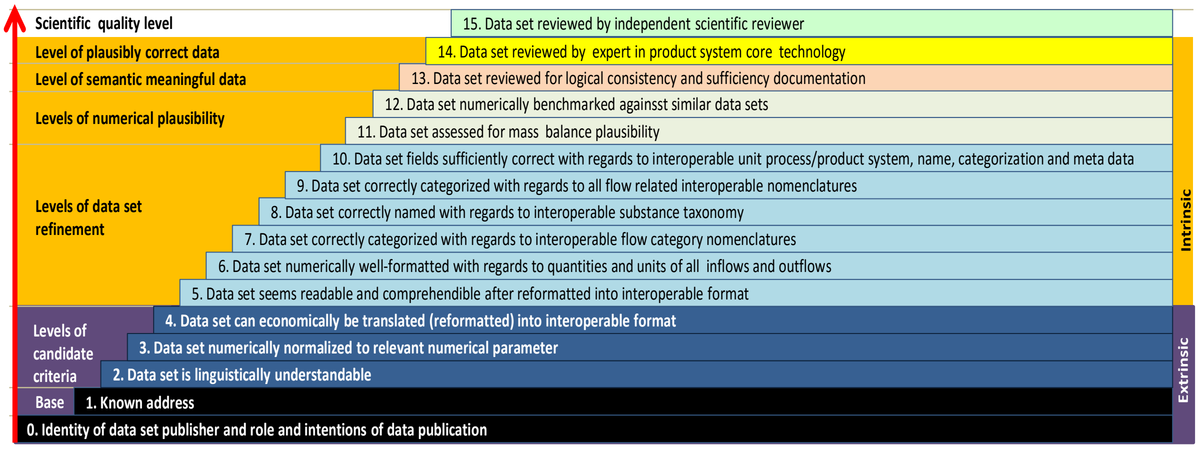


Figure 1: Levels of interoperability (GLAD WG 1 2016, p 3)

While the level 10 definition itself does not help to define or select metadata descriptors[4], it indicates that mass balance checks (level 11) or numerical comparisons (level 12) are out of scope. "interoperable format" (level 4) is relevant for the metadata definition (since metadata is part of a dataset format) but the architecture text does not contain further details about this interoperable format.

1.2. ISO 14040 and 14044

The ISO 14044 series lists ten key descriptors for data quality assessment: time related coverage, geographical coverage, technological coverage, precision, completeness, representativeness, consistency, reproducibility, sources of the data and uncertainty of the information (ISO 14040 2006, ISO 14044 2006). ISO 14044 also requires that the goal and scope of a study be clearly stated and include the intended application, reasons for carrying out the study, the intended audience, and whether the results were intended to be used in a comparative assertion to be disclosed to the public. ISO 14048/TS aims at (ISO 14048/TS, 1 scope)

> "[providing] the requirements and a structure for a data documentation format, to be used for transparent and unambiguous documentation and exchange of Life Cycle Assessment (LCA) and Life Cycle Inventory (LCI) data, thus permitting consistent documentation of data, reporting of data collection, data calculation and data quality, by specifying and structuring relevant information."

It seems therefore an ideal candidate for identifying fields and descriptors for dataset in this report. However, especially related to data quality/fitness for purpose, ISO 14048 foresees only one entry, which is meant to be filled with free text (Figure 2) and one free text entry for the intended use of the dataset.

---

[4] "datasets […] correct with regards to […] interoperable […] meta data" does not provide guidance related to the expected metadata

Refer-ence No.	Data field	Description	Data type	Nomencla-ture	Allowed occur-rences
2.5	**Data quality statement**	Description of known general and specific quality strengths and weaknesses in the process. After having compiled a process, the data generator may be well aware of strengths and weaknesses of the model and the data used to describe it. Such information may however be difficult to detect in the overall documentation of the process. Examples are numerical data that were especially difficult to validate, ambiguity in regard to how to compile an industrial average, or difficulties interpreting data supplied from a site.	Free text	No	One

Figure 2: Data quality in ISO 14048/TS

Since ISO 14044 and ISO 14048/TS deal with similar issues (intended application, data quality), on different levels (dataset, case study), a mapping test between ISO standards was conducted, shown in Annex 12.1. Items in ISO 14044 can often be mapped to ISO 14048/TS, albeit with a rather low level of detail as shown in the test in Annex 12.1. It is worth noting that ISO 14048/TS foresees the assessment of data quality, rather than providing the descriptors which in turn can be used to assess data quality/fitness for purpose. Modeling choices are foreseen to be mentioned in the TS, but limited to the exclusion of processes and flows, and to allocation aspects. This is to some extent certainly due to the age of ISO 14048/TS which was initially released in 2002.

1.3. Recent work on review criteria for LCI datasets

Two consecutive working projects commissioned by SETAC developed, in the last two years, criteria for a review of LCI datasets (Ciroth et al. 2015, Ciroth et al. 2016). The criteria have been presented and discussed in a workshop at the SETAC conference in Nantes in May 2016.

Since a review of a dataset has in the end the aim to investigate the quality or fitness for purpose, the metadata descriptors for datasets should be specified and completed so that they can support a review and the proposed review criteria.

In summary, the working group proposes five main groups of review criteria: goal, model, value, relevance, and procedure:

Goal

- *7 criteria on 'Goal and scope'*
    - Reference time
    - Reference geography
    - Reference technology
    - Reference model completeness
    - Reference sample representativeness
    - Sample approach (scientific, or expert-based)
    - Supported LCIA methods with version number

Model-related criteria, conformance:

These criteria are meant to be assessed for exchanges, and are therefore less relevant for the metadata group. They are:

- *Time related conformance*
- *Geographical conformance*
- *Technological conformance*

- *Model completeness conformance, flows and documentation*

They are of course linked to the goal and scope criteria (time related conformance to reference, for example).

Value: Sample conformance, correctness and reliability

Also these criteria are meant to be assessed on the level of exchanges and can therefore be skipped here, with one exception, the overall consistency of the dataset.

- *Consistency of the provided information*, measured by expert judgement

Relevance, or materiality

- *Mass- and energy balance in line with goal and scope*, as deviation from a complete balance
- *LCIA results in line with goal and scope*, by expert judgement
- *Order of 5 main drivers for main LCIA results in line with goal and scope*, by expert judgement

Procedure

- *Number of reviewers and their relation to the data provider* (dependent or independent, internal or external)
- *Data access* (full, limited, or no access to the supply chain model, or unit process)

2. Structure of the metadata descriptors

2.1. Goal, representation, conformance– the assessment structure

Ultimately, the descriptors are meant to be used in an assessment of fitness for purpose of a dataset. This assessment of fitness for purpose needs to be differentiated in several aspects or topics: the age, the covered technology, modeling details, and so forth. Each of these topics is covered by one metadata descriptor, and for an assessment of each descriptor, three aspects are needed:

- First, a description of the goal for this topic, for the respective dataset. For example, for the topic time, a possible goal can be: Goal for this dataset is to represent the situation for the year 2015. This follows the idea of the Shonan Guidance Principles, where it is proposed to specify and document a goal and scope when modeling a dataset.
- Second, information about what the dataset represents, for the specific descriptor. This information is typically available in the dataset, added by the data provider; we call it therefore **self-declared representation**. In simpler cases, it may be the value of the dataset for a certain descriptor. Again for time, a possible information for representation and value could be: This dataset is representing inputs and outputs valid for 2013.
- Finally and third, a **conformance** assessment complements goal and representation declaration. The conformance assessment shows how far the dataset deviates from the stated goal, for each topic: Taking again time as example, a possible conformance assessment result could be: The dataset is 2 years older than foreseen.

Both representation and conformance assessment result can be reviewed, which is especially useful for self-declared representations.

This structure is in principle valid for any descriptor, and shown in Figure 3.

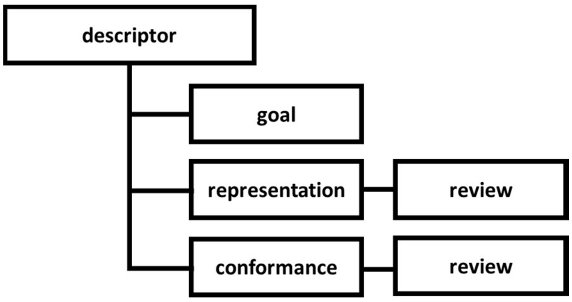


Figure 3: Principal structure of a descriptor; further explanation see text

The application seems straightforward, but gets interesting for two reasons:

1. several existing databases in LCA deal with this concept in a different way. Some store only the representation, some store only the conformance assessment result.
2. for GLAD, it cannot be assumed that users have one uniform purpose / goal and scope when using datasets, therefore a conformance assessment provided by the data provider does not seem appropriate.

For GLAD, it therefore has to be decided, for the identified descriptors, which of the elements of this triple are needed, and where the needed elements are to be stored, in other words how they are to be provided to the user. For the decision, it is useful to distinguish **three types of data descriptors, administrative, flexible, and static**. For static descriptors, the conformance does not depend on goal and scope, because they reflect a common, broadly accepted goal and scope which should not be modified for the datasets in GLAD. Ciroth et al. (2015) have formulated this as follows, for a review of LCA datasets:

> "[…] the dataset should [always and independent from the goal and scope] be
>
> - Correct
> - Understandable and clear
> - Reflect basic LCI dataset structure as described in ISO 14048, with input and output flows and meta information."

Also for GLAD, datasets should always, independent from the user goal and scope, be correct, understandable and clear, and reflect the basic LCI dataset structure. It does not make sense to allow a dataset goal and scope like "this dataset should not be reliable". Therefore, there are descriptors where the conformance assessment will not be depending on goal and scope, and will instead be always identical across different databases; these are called static descriptors. One example is reliability of the source. Flexible descriptors, on the other side, depend on goal and scope.

The differences of these types of descriptors are shown in Figure 4.

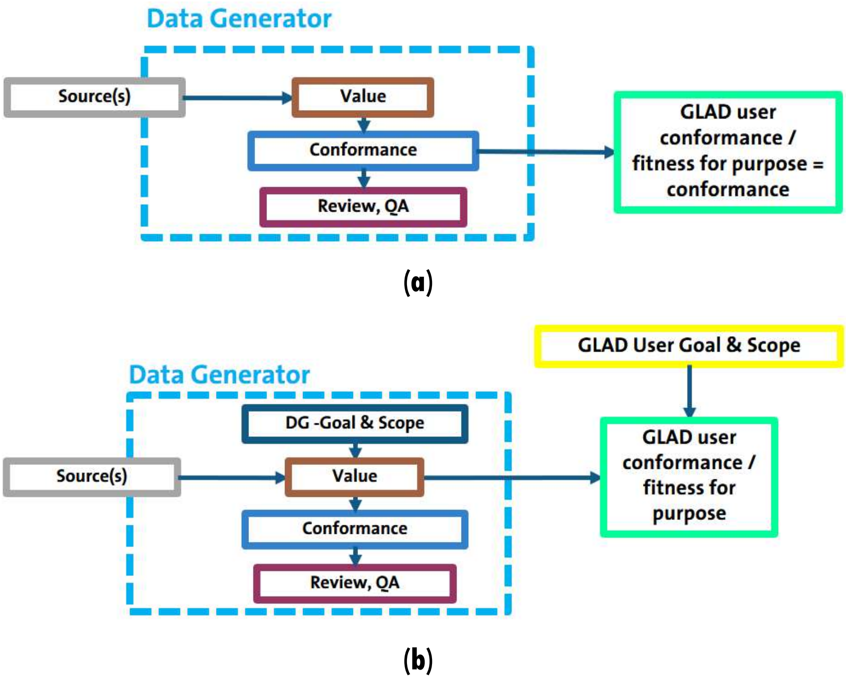
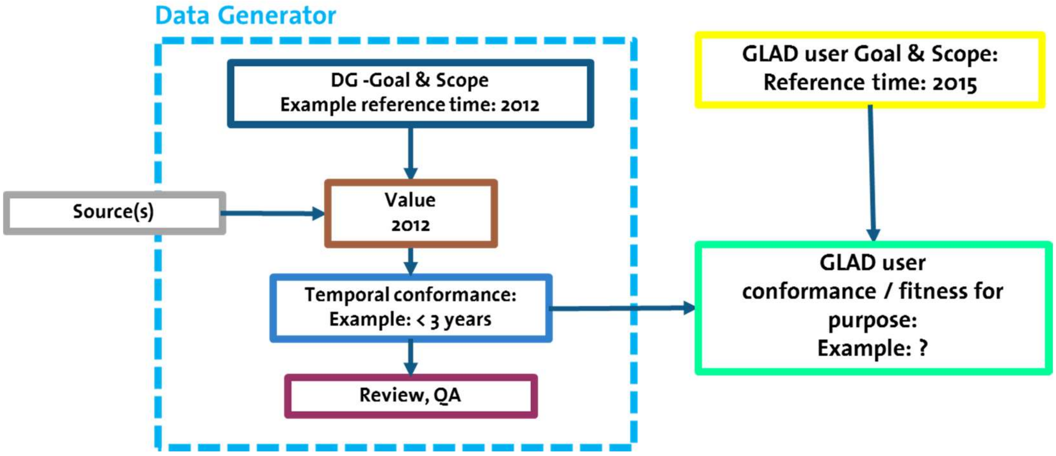


Figure 4: Static data descriptors (a) and flexible data descriptors (b); further explanation see text

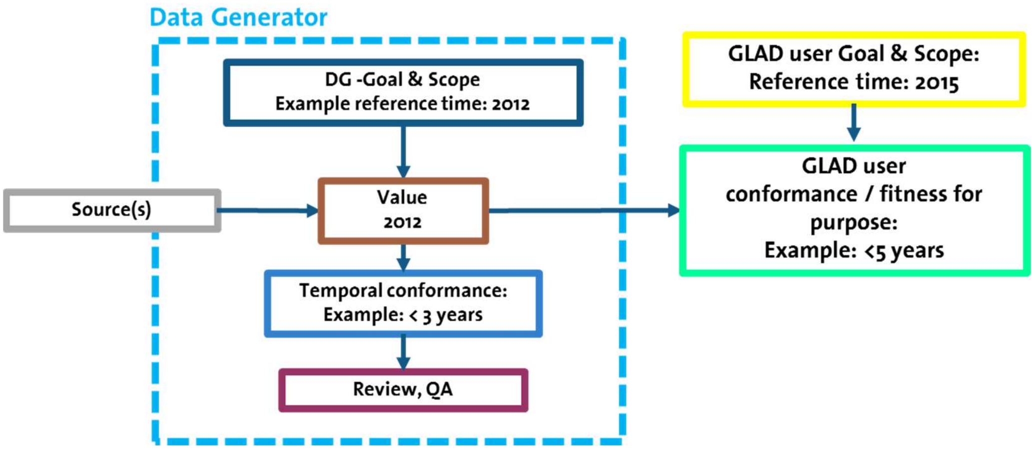
For **static descriptors**, as shown in Figure 4 a, the self-declared representation will not be influenced by dataset and data generator (DG) goal and scope, simply because the goal is static and does not change. Also the GLAD user goal and scope will be identical to the goal and scope of the dataset and thus the assessment of the GLAD user fitness for purpose can be performed with the conformance value alone.

Flexible descriptors, such as temporal, geographical or technological conformance, differ from static descriptors because for these, the goal and scope might be different from dataset to dataset, from data generator to data generator, and from GLAD user to GLAD user. For assessing the GLAD user fitness for purpose, therefore, only the representation is relevant, and not a conformance assessment possibly stored in the dataset, as shown in Figure 4 b.

Looking more closely at storing the various “elements” identified in the figures, it becomes clear that for flexible descriptors, storing the conformance result is less valuable than storing the data representation, the “value”. Figure 5 shows an example of flexible descriptors and the benefit of storing the representation value versus the data conformance. In situation A in Figure 5, only the data conformance is stored, resulting in the inability of a user to properly assess the conformance based on a different goal and scope than the data generator. In situation B, where the time data property is stored, conformance can be continually re-assessed based on the users’ goal and scope.



(A)



(B)

Figure 5: Storing of data property values instead of conformance for flexible descriptors

Administrative descriptors, finally, serve to identify the dataset and are not related to goal and scope.

In conclusion,

- a descriptor has several segments: a value or representation, which is typically self-declared, i.e. provided by the dataset and data provider, a conformance assessment, and a review of the conformance assessment and/or of the representation;
- for a descriptor, not all the segments need to be present, especially the review is optional;
- the conformance assessment depends on goal and scope;
- for a fitness for purpose in the GLAD system, in the end only the GLAD user fitness for purpose / conformance assessment is relevant;
- descriptors can be distinguished into static, flexible, and administrative descriptors; administrative descriptors serve to identify the dataset and are not linked to goal and scope; for static descriptors, the goal and scope will be fix and constant, for flexible descriptors, goal and scope can vary and depend on the application;
- consequently, for static descriptors, the conformance assessment result is more or less equivalent to the self-declared representation, provided the conformance assessment

approach is harmonized across different datasets and data providers, and/or known[5].

- for flexible descriptors, GLAD users will need access to the representation value of datasets, since the GLAD user fitness for purpose determination depends on a varying goal and scope;
- therefore, in GLAD, a metric to assess fitness for purpose / conformance with stated goal and scope will need to be implemented;
- and not the least, users in GLAD will need to be able to specify goal and scope.

2.2. Text fields, language

Many of the descriptor fields are text-based. To support data exchange, it is essential that the text can be understood as broadly as possible. Language barriers can prevent that.

The working group recommends therefore that all descriptors which are text fields are provided in the English language, and optionally in one additional “local” language. The local language is identical for an entire process dataset and the related flow, flow property, unit datasets. The local language needs to be specified.

This proposal follows the specification of the old EcoSpold01 format, where German always was the local language; it is more rigid than e.g. the ILCD format where any text field in one dataset can be provided in any language, with a possible outcome of a dataset with text in Hungarian, English, German, and Japanese altogether.

name	401	Text	80
	Name of the unit process, elementary flow or impact category.		
	For unit processes and system terminated name is used as the identifying product/service, production process or worked product, level of processin (at plant, at regional storehouse), or destination (for wastes: to sanitary la		
	For elementary flows name, unit, category and subCategory are used as t		
	For impact categories, name, location, unit, category and subCategory are		
	English is the default language in the ecoinvent quality network.		
localName	490	Text	80
	see 'name' for explanations.		
	German is the default local language in the ecoinvent quality network.		

Figure 6: Excerpt from the EcoSpold01 schema documentation, process: name and localName for a text field

3. Metadata descriptor list and shortlist

The scope of this task of the metadata working group is to define a set of metadata descriptors, which is essential for a user to determine “fitness for purpose.” It is outside the scope of this work to consider

1) additional metadata descriptors necessary for search/data discovery, and
2) specific field formats or data types for each of the descriptors.

This limitation of the focus is done for practical reasons. It was often stressed that the purpose of the group is further to define a “minimum set” or shortlist of descriptors.

The full metadata descriptor list is presented in the following, initially without specifying the format, but with indication of the descriptor segment (see section 2.1). Most of the descriptors are for assessing fitness for purpose in relation to goal and scope of the dataset. A very basic set of descriptors are related to a clear identification of the dataset. These descriptors are seen as

[5] Some information might be lost from representation to conformance assessment, for example if the conformance assessment classifies the representation in ordinal scales.

mandatory. Descriptors essential for understanding fitness for purpose of a dataset are proposed mandatory and recommended. Further descriptors exist and are seen as optional. The shortlist are those descriptors which are mandatory or recommended. Section 7.3 provides a motivation and discussion for the short list.

The complete set of descriptors is provided in Figure 7, structured into

- Identification (ID)
- Descriptors for which goal, value/representation, and conformance are proposed (gvc descriptors)
- Descriptors related to modeling (modeling)
- Descriptors related to sampling (sampling) and to calculation (calculation)
- Descriptors related to quality assurance (QA), and finally,
- Descriptors related to administrative aspects (administrative)

In Figure 7 and in the following discussion, the elements of the descriptors (see 2.1) are listed separately. This is in accordance with current practice in LCA and LCA data formats, where often only one of the elements of a descriptor is provided. For clear identification, the elements are numbered, with “I” for goal, “II” for conformance, and other numbers for value and representation (0, III, IV, V, V, VII). For some descriptors, only value is relevant (e.g., dataset contact). For other descriptors, only one element, typically the value, is proposed, although also here goal and conformance assessment can be imagined (Figure 7).

		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		IIIa	
Calculation	Aggregation type if any		VIa	
QA	Data set review performed		Va	
	Type of data set review		Vb	
	Quality assurance performed		Vc	
	Reviewing person(s)		Vd	
Administrative	Copyright protected data set?		VIIa	
	Copyright holder		VIIb	
	Free data set or for purchase?		VIIc	
	Data set license		VIId	
	Data set contact		VIIe	

Figure 7: Summary of the proposed descriptors, for further explanation see text

The descriptors which are proposed to be created in GLAD can either be created on the fly, upon user demand, or rather pre-calculated and stored in GLAD. This can be decided in a later implementation.

Table 1: List of proposed metadata descriptors

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Necessity ***	Comment / Example
0	Dataset identification							
a	Process name	General descriptive name of the process and/or its main good(s) or service(s) and/or its level of processing. The dataset name ideally conforms to a harmonized nomenclature convention which could be specified in the GLAD nomenclature group. An example could be: Base name; treatment, standards, routes; mix and or location type; quantitative product or process properties	Dataset	Ad	DaP	Value	M	Example: corn grain; average tillage practice mix; at farm; 15% moisture
b	Process type	Unit process or aggregated process	Dataset	Ad	DaP	Value	M	
I	Goal and scope							
a	Reference time	Point in time or time period foreseen for the dataset	Dataset	FI	GLAD	Goal	O	
b	Reference geography	Reference area or point foreseen for the dataset	Dataset	FI	GLAD	Goal	O	
c	Reference technology	Technology or technology mix foreseen for the dataset.	Dataset	FI	GLAD	Goal	O	
d	Reference model completeness	Targeted coverage of inputs and outputs in the dataset	Dataset	FI	DaP	Goal	PM	In terms of numbers of flows and /or a complete coverage of the amounts, leading to a complete mass and / or energy balance
e	Reference sample representativeness	Targeted sample representativeness, in line with the	Dataset	FI	DaP	Goal	PM	The intended sampling approach is If)

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Necessity ***	Comment / Example
		foreseen sampling approach						
f	Intended sample approach	Scientific or expert-based sampling	Dataset	FI	DaP	Goal	PM	Or: expert judgement as default
g	LCIA methods with version number to be supported	See name	Dataset	FI	DaP	Goal	R	CML baseline, Version 4.4. of January 2015
II	Conformance							
a	Time related conformance	Time difference between the reference time foreseen for the dataset, according to documentation, and the time period for which data were surveyed (i.e. the period of the initial data collection).	Exchange	FI	GLAD	Conformance	R	
b	Geographical conformance	Geographical difference between the reference area foreseen for the dataset, according to documentation, and the area for which data were surveyed	Exchange	FI	GLAD	Conformance	R	
c	Technological conformance	Technological difference between the reference technology or technology mix foreseen for the dataset, according to documentation, and the technology for which data were surveyed.	Exchange	FI	GLAD	Conformance	R	
d	Model completeness conformance	Coverage of inputs and outputs in the dataset in relation to the targeted coverage	Dataset	FI	DaP	Conformance	R	
e	Sample conformance	Representativeness of the information provided, in relation to the sample conformance specified	Exchange	FI	DaP	Conformance	R	Assessed according to the specified sample approach

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Necessity ***	Comment / Example
		in goal and scope						
III	Reliability of the sources used							
a	Reliability of the sources	Reliability of the information provided in the dataset, assessed by reliability of the sources used for obtaining the information	Exchang e	st	DaP	Cross-cutting	R	
IV	Life cycle and model							
a	Time	Time or time period relevant	Exchang e	FI	DaP	Value	PM	
b	Geography	Geographical area or point relevant	Exchang e	FI	DaP	Value	PM	
c	Technology	Technology or technology mix relevant	Exchang e	FI	DaP	Value	PM	
d	Supported LCA nomenclature system(s)	See name	Dataset	FI	DaP	Value	R	ILCD reference flow list, ILCD 1.1 from May 2015; Ecoinvent 3.3 master data
e	Supported LCIA methods	See name	Dataset	FI	DaP	Value	R	CML baseline, Version 4.4. of January 2015
f	Materiality, completeness	Order of 5 main drivers for main LCIA results	Dataset	FI	DaP	Value	R	LCIA categories according to goal and scope; drivers are exchanges / elementary flows for unit processes, aggregated processes and elementary flows for aggregated processes
g	Representativeness	For science-based sampling, variation coefficient plus documentation; for expert judgement, representativeness classes estimates	Exchang e	St	DaP	Value	PM	Static descriptor since it seems always desirable to have a representative dataset

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Necessity ***	Comment / Example
h	LCI modeling approach	Attributional or consequential modeling	Dataset	FI	DaP	Value	PM	See also detailed format description, chapter 4
i	System boundaries	Set of criteria specifying which unit processes are part of a product system (ISO, 2006)	Dataset	FI	DaP	Value	R	
j	Method used to deal with multifunctional processes	Method used to assign the environmental burdens to the joint production of the reference flows.	Dataset	FI	DaP	Value	PM	
k	Biogenic carbon	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).	Dataset	FI	DaP	Value	PM	
l	Land use	See detailed format description, chapter 4	Dataset	FI	DaP	Value	R	
m	Wastes and end-of-life	Substances or objects which the holder intends or is required to dispose of (ISO 14040, 2006) .	Dataset	FI	DaP	Value	R	
n	Water	Water use	Dataset	FI	DaP	Value	R	
o	Infrastructure/capital goods	Product not intended for consumption, with a lifetime exceeding one year (Weidema et al., 2013).	Dataset	FI	DaP	Value	R	
p	Long-term emissions (beyond 100 years)	Emissions that will occur in the future but are determined today (EC-JRC, 2010).	Dataset	FI	DaP	Value	R	
q	Temporary carbon storage, delayed greenhouse gas emissions, delayed credits for solving multifunctionality	See detailed format description, chapter 4	Dataset	FI	DaP	Value	R	
V	Verification and quality assurance							

#	Metadata descriptor name	Definition	Scope	Type*	Data created by**	Descriptor segment	Necessity ***	Comment / Example
a	Dataset review performed	See name	Dataset	Fl	DaP	Value	PM	
b	Type of dataset review	Number of reviewers and their relation to the data provider, internal or external	Dataset	Fl	DaP	Value	R	
c	Quality assurance performed	Independent quality assurance related to the information provided, on single entries of the dataset	Exchanges & dataset fields	Fl	DaP	Value	R	
d	Reviewing person(s)	Person who conducted the review	Dataset	Fl	DaP	Value	PM	
VI	Calculation							
a	Aggregation type if any	For an aggregated dataset, specify how the aggregation was performed.	Dataset	Fl	DaP	Value	If 0b) ≠ unit process: PM	Horizontal, along the supply chain / vertical, across several processes delivering similar products, mixed; partial or complete – so, 6 cases)
VII	Administrative							
a	Copyright protected dataset?	See name	Dataset	Ad	DaP	Value	M	
b	Copyright holder	See name	Dataset	Ad	DaP	Value	If VIIa) = true: M	
c	Free dataset or for purchase?	See name	Dataset	Ad	DaP	Value	If VIIa) = true: M	
d	Dataset license	See name	Dataset	Ad	DaP	Value	If VIIa) = true: M	
e	Dataset contact	See name	Dataset	Ad	DaP	Value	M	

*Ad: administrative; St: static; Fl: flexible
**DaP: data provider

***O: Optional, user-defined; R: recommended; M: mandatory; PM: proposed mandatory

4. Format recommendations and requirements for the metadata descriptors

Building on the previous chapter, for the metadata descriptors proposed, more detailed information is provided in the following. Since the additional information for all the descriptors follows a common pattern, the information is provided in the form of tables.

The additional information includes:

- A more detailed definition (for some descriptors)
- Indication towards a format; this is not meant to be a full format specification, but provided with the aim to support a future implementation of the descriptor, in software and databases, and obviously following an LCA data format
- A cross-reference to existing implementations, for some descriptors

For the text fields, see also the remark on languages in section 2.2.

This initial format recommendation is seen as input for a future implementation and also as input for a more in-depth discussion with existing format "operators" (i.e. organizations who are releasing data formats) and not the least data providers.

<u>0 Dataset information</u>

a. Name

Background	
Definition	General descriptive name of the process and/or its main good(s) or service(s) and/or its level of processing. The dataset name shall conform to ILCD convention (EC-JRC, 2010): Base name; treatment, standards, routes; mix and or location type; quantitative product or process properties Example: corn grain; average tillage practice mix; at farm; 15% moisture
Scope	Dataset

b. Process type

Background	Although datasets are preferable as unit process in order to facilitate the reviewing process, increase the transparency, allows flexibility and adaptation by users and improve the interpretation of an LCA study, there are also reasons to aggregate datasets, as follow: to reduce calculation times and due to confidentiality issues (UNEP, 2011). Despite the aggregation can be distinguished between two types (vertical aggregation and horizontal averaging), this descriptor refers only to the vertical aggregation. This seems reasonable since horizontal aggregation of a single process can also be called a unit process. In addition, a vertical aggregated process can be fully terminated (if will be made only of elementary flows) or partially terminated (if will stand as a mix of both elementary and non-elementary flows) (UNEP, 2011).
Definition	Unit process: smallest element considered in the LCI analysis for which input and output data are quantified (ISO, 2006). Aggregated process: the combination of unit processes that succeed each other in a product life cycle, connected with intermediary flows (UNEP, 2011, p.68).
Scope	Dataset

I Goal and scope

a. Reference time

Background	The reference time is expected to influence the relevance of a particular dataset to a particular use because it affects the technology level. Technologies change over time, the efficiency of the production processes increases, and what was the average technologies years ago now could contribute to the market mix only to a limit extent. The reference time can be a certain point in time, a part of a year, an entire year, or several years. Primary technology is considered under technological conformance so should not be considered here.
Definition	Reference time foreseen for the dataset.
Scope	Dataset

b. Reference geography

Background	Geography is expected to influence the relevance of a particular dataset to a particular use because it affects "production conditions", where these include the primary technologies used, regulatory standards, technologies used for second-order activities (e.g ancillary, transportation activities) climatic conditions, water availability, density of land use, and soil type. Primary technology is considered under technological conformance so should not be considered here.
Definition	Reference area foreseen for the dataset
Scope	Dataset

c. Reference technology

Background	The technology of the product or service to be addressed in the dataset.
Definition	Technology or technology mix foreseen for the dataset.
Scope	Dataset

d. Reference model completeness

Background	The idea is to guarantee that the dataset targeted coverage regarding to all the key processes, parameters and elementary flows will comply with the overall requirements established in the goal definition and intended applications of the dataset - all non-reference product flows, waste flows and elementary flows that are quantitatively irrelevant can be ignored; they can be "cut-off" (Frischknecht et al. 2007, p 10; EC 2010b, p 99) in UNEP (2011, p.62). The information on the intended model completeness support the user's decision if the dataset can be used as proxy background data for some inputs/outputs on the user's system model (e.g.in case that the dataset has low completeness but did not represent a relevant input/output of the user's model) or can be used as reference for system comparison (e.g. in case that the dataset has high completeness). The model completeness encompasses all model-related aspects, which includes the reference flows used, and other dataset modelling aspects (i.e. IVa) to IVq) in Table 1) as well (Ciroth et al., 2016). If deficiencies are noted, additional efforts are required to fill the gaps. In some cases, it may be possible to close data gaps – in these instances, the differences in the data should be

	reported in documentation (USEPA 2006) in UNEP (2011, p.62)
Definition	The agreement whether the modelling and data from a dataset is in accordance with the goal and scope definition (intended application) of the dataset.
Scope	Dataset

e. Reference sample representativeness

Background	When collecting data for a dataset, often one will not be able to gather all the data required to generalize the results to represent the intended population, this can also be called the actual sample. The reference sample representativeness, on the other hand, refers to the intended sample that should be used to fulfil the requirements of completeness. This descriptor would be to define the reference sample representativeness, while the conformance descriptors (i.e. IIe) in Table 1) will evaluate if the actual sample is in conformance with sample representativeness.
Definition	Targeted sample representativeness, in line with the foreseen sampling approach
Scope	Dataset

f. Intended sample approach (scientific, or expert-based)

Background	Although a “scientific” – i.e. statistically robust – approach to sampling populations of processes to obtain LCA data is the ideal approach, in practice it is not always used or feasible (Ciroth et al., 2016). Therefore, often sample data are based on expert judgement.
Definition	Method used to define a value (it can be a process, parameter or flow)
Scope	

g. LCIA methods with version number intended to be supported

Background	LCA studies shall cover an impact assessment in the three areas of protection. There are a wide range of LCIA methods that allow an impact assessment at midpoint and endpoint level, moreover, each LCIA methodology has different number of impact/damage categories and follows a specific method for each impact/damage category covering different substances (elementary flows) and spatial and temporal conditions. Ideally, a dataset should cover all the potential impacts, however, this is not always possible, and therefore the dataset must clearly define the impact categories (e.g. Water depletion, Carbon footprint, etc.) that can be covered for an impact assessment and for which methodology (e.g. Pfister et al., IPCC (2007), among others).
Definition	LCIA methods that are planned to be supported by the dataset. Ideally, LCIA methods should be easy to identify and follow a nomenclature
Scope	Dataset

<u>II Conformance</u>

a. Time related conformance

See section 3.

b. Geographical conformance

See section 3.

c. Technological conformance

See section 3.

d. Model completeness conformance

See section 3.

e. Sample conformance

See section 3.

III Reliability of the sources used

See section 3.

IV Life cycle and model

a. Time

See section 3.

b. Geography

See section 3.

c. Technology

See section 3.

d. Supported LCA nomenclature systems

Background	Different LCA working groups use often considerably different nomenclature (EC-JRC, 2010). In consequence, LCI datasets are incompatible on different levels, what strongly limits the combined use of LCI datasets, data exchange and the impact assessment (EC-JRC, 2010).
Definition	Nomenclature is a set of rules to name and classify data in a consistent and unique way (ISO, 2002).
Scope	Dataset

e. Supported LCIA methods

Background	While Ig) addresses the LCIA method the dataset should support (goal), here, the LCIA method that actually are supported are documented
Definition	LCIA methods supported by the dataset, especially regarding the elementary flows and compartments provided with the dataset, see also Ig). Ideally, LCIA methods should be easy to identify and follow a nomenclature
Scope	Dataset

f. Materiality, completeness

Background	The idea here is to identify whether the dataset is able to provide information on the main drivers for LCIA results, meaning elementary flows, and for processes in case the dataset is an aggregated dataset. The selected LCIA method must be in line with goal and scope. “Main results” can be identified either by the reviewer, using expert judgement, or can be specified in goal and scope. Also whether the LCIA results are in line with goal and scope

	needs to be assessed by the reviewer.
Definition	Order of 5 main drivers for main LCIA results in line with goal and scope
Scope	Dataset

g. Representativeness

Background	“For an LCA dataset, information will typically vary by technology and producer and region, and by time. This criterion assesses how representative the information regarding these four aspects is, for the dataset. Aim for a perfect score is information that is representative in a statistical sense. Simply speaking, this is fulfilled if the information is obtained by random sampling, i.e. a sampling where all items of interest, called “population” in statistics, have a known chance of being drawn (e.g., Hansen et al. 1953, Vol I p 9). This sampling is not common practice in LCA, although some few examples exist. One example for LCA is described in (Ciroth et al. 2008) for packaging, another is presented in (Yodkhum, Sampattagul 2014) for rice production on paddy fields in Thailand. Therefore, in goal and scope, the foreseen sampling approach is to be noted (see goal and scope). The assessment of this criterion assumes by default, that very good sample conformance is aimed for. If this does not hold for a dataset, then this needs to be mentioned in goal and scope, and the default scores provided below need to be adjusted accordingly. Measure for sample conformance is for science based sampling the coefficient of variation of the sample, while for expert-based sampling, the proposed measure is expert judgement. It is not recommended to use, for expert-based sampling, parameters such as market share or similar to guess the representativeness of the sample since these are misleading” (Ciroth et al. 2016, sample conformance) For science-based sampling, the variation coefficient is used, as dimensionless measure for the error in the sample; for expert-based sampling, it is questionable whether any key parameters such as market share or similar can be reasonably used to estimate the sample representativeness; therefore, “pure” expert judgement is used instead.
Definition	Representativeness of the information provided
Scope	Exchange and dataset field

h. LCI modelling approach

Background	According to the goal and scope of the LCI dataset it can be applied two modelling approach: the attributional and the consequential system modelling. The choice between these two different approaches, besides answering different research questions, has implications for other later choices such as input data and how the multifunctionality problems must be solved.
Definition	Attributional modeling is defined as system modelling approach in which inputs and outputs are attributed to the functional unit of a product system by linking and/or partitioning the unit processes of the system according to a normative rule (UNEP, 2011). Consequential modeling is defined as system modelling approach in which activities in a product system are linked so that activities are included in the product system to the extent that they are expected to change as a consequence of a change in demand for the functional unit (UNEP, 2011). Consequential modeling is hard to describe in short descriptor values. We propose it here as a descriptor value in order to flag

	those datasets where consequential modeling is used, so that any user of the dataset can check the dataset documentation for further details
Scope	Dataset

i. System boundaries

Background	The system boundary determines, for aggregated processes, which processes of a system are included/excluded in the dataset, typically the location of the start and end of the process seen as a chain of activities. For a unit process, the system boundary determines what is included in the dataset. The criteria and rationale of the cut-off rules for the exclusion of specific flows must be described, i.e. low mass, energy and environmental relevance or lack of information. This descriptor is linked to the model completeness and also to the materiality descriptor.
Definition	Set of criteria specifying which unit processes are part of a product system (ISO, 2006)
Scope	Dataset

j. Method used to deal with multifunctional processes

Background	The use of mathematical artifices to deal with multifunctional process has long been discussed by the LCA community since the process subdivision is not always possible and system expansion is not feasible once it can lead to the need for further system expansion. To the latter, the so-called substitution method is generally applied as a way of system expansion. The consistent use of the substitution method must be clearly described regarding to the marginal product/technology replaced. Whenever the system expansion (substitution method) is not feasible, ISO recommends the use of partitioning methods based on causal relationship of the input and output flows, the physical relationships of the references flows and market prices.
Definition	Method used to assign the environmental burdens to the joint production of the reference flows.
Scope	Dataset

k. Biogenic carbon

Background	Biogenic carbon (and biogenic methane) is the emission related to the natural carbon cycle and those from crops, animal husbandry and biobased products. It is important to differentiate the carbon type when (possibly) not all biogenic carbon emissions are reported or when biogenic flows are given different importance in the LCIA.
Definition	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).
Note on implementation	A Boolean (yes/no) field is not sufficient since there are several possible approaches for dealing with biogenic carbon. Therefore an enumeration of possible approaches is proposed, with one “other” option, and an additional text field for adding explanation.
Scope	Exchange

l. Land use

Background	Despite the use of land by the product system (calculated by the area in a certain time), the impacts from natural land transformation should be inventoried and distinguished in direct (dLUC) and indirect (iLUC) change. Land use change implies in CO2 emissions due to soil carbon changes, aboveground biomass, belowground biomass

	and dead organic matter. Ideally, the methods and tools used to estimate the land transformation and the carbon emissions must be described (i.e. emissions factors, time period, etc.).
Definition	Land occupation (also called land use) occurs when the occupied land is prevented from changing to a more natural state i.e. the properties of a piece of land are maintained (Allacker et al., 2014; Koellner et al., 2013). Land transformation (also called land use change) occurs when there is change in the land use type i.e. the properties of a piece of land are modified (Allacker et al., 2014; Koellner et al., 2013).
Scope	Exchange

m. Wastes and end-of-life

Background	Waste flows are in LCA processes typically “the opposite” of product flows; while product flows are produced in one process and used in several other processes potentially, a waste flow is created in several different processes, and taken up by one waste treatment process potentially. Waste flows can be differentiated with positive market value and negative market value. For the former an approach to deal the multifunctionality must be defined. While for the latter, for LCA all inputs and outputs should be modelled until the inventories exclusively shows elementary flows (EC-JRC, 2010). For waste with positive market value different approaches can be applied to deal with multifunctionality, such as cut-off, allocation (physical, economic) in the point of substitution, substitution approach, with implications on the environmental burdens related to the product system that generate the waste and to the product system that uses the treated waste as secondary material.
Definition	Substances or objects which the holder intends or is required to dispose of (ISO 14040, 2006) .
Note on implementation	A Boolean (yes/no) field is not sufficient since there are several possible approaches for dealing with, and for defining, waste flows. Therefore, in each case, an enumeration with possible approaches is proposed, with one “other” option, and an additional text field for adding explanation.
Scope	Dataset

n. Water

Background	This descriptor should describe the method used to calculate water consumption in crop production and the water types (e.g. water withdrawal, green water), sources (e.g. river, groundwater) and the degradative water use (i.e. on the output side) (e.g. wastewater, emission in form of steam).
Definition	Water use.
Scope	Exchange

o. Infrastructure/capital goods

Background	Infrastructure although frequently not consider in LCA modelling can show be very relevant for certain sectors and or processes with fewer direct emissions during operation but with material-intensive infrastructure (GaBi 2014; Frischknecht et al., 2007).
Definition	Product not intended for consumption, with a lifetime exceeding one year (Weidema et al., 2013).
Scope	Exchange

p. Long-term emissions (beyond 100 years)

Background	The logic for separation of short-term and long-term emissions is that both have different uncertainty: emissions today can be measured, emissions from beyond 100 years can only be roughly forecasted (EC-JRC, 2010). Moreover, in LCA, the use of a 100-year time horizon for assessing global warming impacts implies a cut-off of the 'tails' of GHG's atmospheric residences at 100 years following their emission (Brandão et al., 2013). Ideally, the LCI should consider both emissions separately.
Definition	Emissions that will occur in the future but are determined today (EC-JRC, 2010).
Note on implementation	A Boolean (yes/no) field is not sufficient since there are several possible approaches for dealing with long-term emissions. Therefore, in each case, an enumeration with possible approaches is proposed, with one "other" option, and an additional text field for adding explanation.
Scope	Exchange

q. Temporary carbon storage, delayed greenhouse gas emissions, delayed credits for solving multifunctionality

Background	Carbon sequestration and temporary storage is often discussed as a means to mitigate climate change, whether a bio-based or fossil-based product or process, the latter due to the delayed carbon emissions (Brandão et al., 2013). Although there are significant efforts to develop robust methods to account for these benefits, there still no consensus on the most appropriate way of considering the sequestration, temporary storage and release emissions (Brandão et al., 2013).
Definition	Carbon sequestration refers to the removal of carbon dioxide from the atmosphere (Brandão et al., 2013). Temporary storage refers to the maintenance of the sequestered carbon for a limited period of time in non-atmospheric pools (Brandão et al., 2013).
Scope	Exchange

V Verification and quality assurance

a. Dataset review performed

Background	A dataset that has successfully passed a review can typically be used with less own quality assurance and refactoring required, given the review goal and scope is comparable to own goal and scope.
Definition	Has the dataset successfully passed a review in the LCA context
Scope	Dataset

b. Type of dataset review, number of reviewers

Background	This descriptor is only relevant once Va) is completed, i.e. for datasets which have passed a review
Definition	Number and type of reviewers
Scope	Dataset
Format	1 One or more independent external reviewer 2 Two or more independent internal reviewers 3 One independent internal or two or more internal reviewers 4 One internal reviewer

c. Quality assurance performed

Background	Apart from reviewing the entire dataset, it can make sense to perform a quality assurance for single entries and fields in the dataset. This descriptor has the purpose to document these. The descriptor can also be applied for a group of similar dataset fields or exchanges.
Definition	Quality assurance and verification method applied, and result.
Scope	Exchange and dataset field

d. Reviewing person(s)

Background	Persons responsible for the review should be added.
Definition	Contact details to persons responsible for the review (if performed)
Scope	Dataset

VI Calculation

a. aggregation type if any

See section 4.

VII Administrative

a. copyright protected dataset?

See section 4.

b. Copyright holder

See section 4.

c. free dataset or for purchase?

See section 4.

d. dataset contact

See section 4.

5. Two test case applications

To demonstrate and test the proposed descriptors, they will be applied for two different test cases, from the ELCD network and from ecoinvent. Both datasets are also available in LCA software; however, for the test, the datasets are used as they are available "from the source", i.e. from the ecoinvent website in EcoSPold02 format, and from the ELCD network in ILCD format.

Each case is explored in a feasibility test, where it is checked how well the foreseen metadata descriptors are supported by the test dataset, and in a GLAD user mock test, where an assumed GLAD user assesses fitness for purpose of the test dataset.

Both cases were randomly selected. They can obviously not claim to be fully representative of the two considered data sources, but follow a case study approach: a specific, non-representative case is investigated thoroughly, in order to find out a real situation as closely as possible, and with as little "methodological overhead" as possible as well (Yin 2003).

5.1. Case 1: Computer production, laptop, global region, from ecoinvent 3.3

The first dataset is a dataset from ecoinvent, a laptop computer. It was already present in the ecoinvent 2 database and has been migrated by ecoinvent to the EcoSpold02 format, it is now integrated in the ecoinvent 3.3 database.

Process dataset UUID: b525f1a4-59ef-4da0-b6e1-f188c4358881.

5.1.1. Test for GLAD descriptor support

Overview

For the test, the original dataset in EcoSpold02 format is compared against the proposed metadata descriptors; the EcoSpold02 XML structure is shown with a

```
Brown background.
```

The descriptors are shown in a table. For documenting the assessment, a color code is used, as follows:

A green background is provided if the descriptor is supported by the dataset
A green background with red font is used to show limitations for a descriptor which is generally supported by the dataset
An orange background is used of the descriptor is not supported by the dataset
A yellow background is used for aspects that could not be tested or that are not fully supported
A light grey background finally is used for the topic areas of the descriptors

The test goes through all the descriptors proposed, as listed in Table 1. Only those descriptors are addressed which are to be provided by the data provider. Focus is to investigate whether the information is available, rather than the format of the information.

Results for the descriptors

0 Dataset information

Dataset information can be easily found in the dataset, as expected.

```
<ecoSpold xmlns="http://www.EcoInvent.org/EcoSpold02">
  <activityDataset>
    <activityDescription>
      <activity id="b525f1a4-59ef-4da0-b6e1-f188c4358881" activityNameId="1a8bda5d-
80e1-4ecc-9ff6-7b898b5ec8a2" inheritanceDepth="0" type="1" specialActivityType="0"
energyValues="0">
        <activityName xml:lang="en">computer production, laptop</activityName>
        <synonym xml:lang="en">Notebook, Lifebook</synonym>
```

a	Process name	computer production, laptop
b	Process type	1 (unit process)

I Goal and scope

Goal and scope for the dataset is not available directly attached to the dataset but it rather needs to be extracted from a methodological background report; in the case of the specific dataset, even two reports are relevant since the dataset has been created in a previous version of ecoinvent and was migrated to the more recent version.

```
intermediateExchangeId="14a9fc00-5a28-40f3-9cbe-527a40212577">
        <name xml:lang="en">polystyrene, high impact</name>
        <unitName xml:lang="en">kg</unitName>
        <comment>EcoSpold01Location=RER</comment>
        <comment xml:lang="en">Literature Value.</comment>
        <uncertainty>
          <lognormal meanValue="0.42395" mu="-0.86" variance="0.02"
varianceWithPedigreeUncertainty="0.0301" />
          <pedigreeMatrix reliability="1" completeness="4" temporalCorrelation="1"
geographicalCorrelation="3" furtherTechnologyCorrelation="3" />
        </uncertainty>
```

d	Reference model completeness	Targeted coverage of inputs and outputs in the dataset	Not explicitly provided in the dataset nor in the methodological background report; implicitly, goal seems to include all relevant environmental impacts ("environmental knowledge of the people involved in compiling the LCI data is used to judge whether or not to include a certain input", Figure 8)
e	Reference sample representativeness	Targeted sample representativeness, in line with the foreseen sampling approach	Not explicitly provided in the dataset; in the methodological background report, for the pedigree matrix, the ideal assessment of reference sample representativeness asks for "representative data from all sites", Figure 9, score 1
f	Intended sample approach	Scientific or expert-based sampling	Not explicitly provided in the dataset; in the methodological background report, for the pedigree matrix, the foreseen assessment of reference sample representativeness follows an expert-estimate approach (with market share percentages as indicator for representativeness), Figure 9
g	LCIA methods with version number to be supported	See name	Not explicitly provided in the dataset nor in the methodological background report

4.5.2 Cut off rules

The analyses of technical processes required to manufacture products and deliver services are based on pure environmental process chain analysis. Results from enlarged economic input-output analyses are only used in exceptional cases. Hence, data shown in the reports and the ecoinvent database are based on process life cycle inventories (and neither input-output nor hybrid life cycle inventories).

According to ISO 14044 (International Organization for Standardization (ISO) 2006b) several criteria are used to decide which inputs to be studied, including a) mass, b) energy, and c) environmental relevance.

No strict quantitative cut-off rule is followed in the ecoinvent project. Environmental knowledge of the people involved in compiling LCI data is used to judge whether or not to include the production of a certain input or whether or not to include the release of a certain pollutant.

Sometimes it can be more relevant to investigate process specific emissions rather than the energy requirements because the latter may be of only little environmental relevance. Specific environmentally relevant issues (e.g., the emission of a particular pollutant) shall be included at the cost of the completeness of a unit process' mass and energy flows.

Figure 8: Cut-off rules for ecoinvent 2 (Frischknecht, Jungbluth et al. 2007), for Id, reference model completeness

Indicator score	**1**	**2**	**3**	**4**	**5** (default)
Completeness	Representative data from all sites relevant for the market considered, over an adequate period to even out normal fluctuations	Representative data from >50% of the sites relevant for the market considered, over an adequate period to even out normal fluctuations	Representative data from only some sites (<<50%) relevant for the market considered ***or*** >50% of sites but from shorter periods	Representative data from only one site relevant for the market considered ***or*** some sites but from shorter periods	Representativeness unknown or data from a small number of sites ***and*** from shorter periods

Figure 9: Pedigree matrix for ecoinvent 3, completeness, (Weidema et al. 2013, p 76), for Ie, reference sample representativeness

II Conformance

Conformance obviously depends on goal and scope; since model completeness goal and scope is not provided in an "operational" way for the dataset, the conformance cannot be assessed either.

d	Model completeness conformance	Coverage of inputs and outputs in the dataset in relation to the targeted coverage	Since no explicit model completeness goal and scope is formulated for the dataset (see I d), model completeness conformance is difficult to assess
e	Sample conformance	Representativeness of the information provided, in relation to the sample conformance specified in goal and scope	Provided for each exchange (as foreseen for GLAD), see the example above, for one of the exchanges: `completeness="4"`

III Reliability of the sources used

Reliability of the sources is available for each exchange in the dataset. The information is only available as conformance assessment (pedigree result), but, since the descriptor is static, the conformance assessment is equivalent to a value and representation and can be used as well.

```
intermediateExchangeId="14a9fc00-5a28-40f3-9cbe-527a40212577">
        <name xml:lang="en">polystyrene, high impact</name>
        <unitName xml:lang="en">kg</unitName>
        <comment>EcoSpold01Location=RER</comment>
        <comment xml:lang="en">Literature Value.</comment>
        <uncertainty>
          <lognormal meanValue="0.42395" mu="-0.86" variance="0.02"
varianceWithPedigreeUncertainty="0.0301" />
          <pedigreeMatrix reliability="1" completeness="4" temporalCorrelation="1"
geographicalCorrelation="3" furtherTechnologyCorrelation="3" />
        </uncertainty>
```

a	Reliability of the sources	Reliability of the information provided in the dataset, assessed by reliability of the sources used for obtaining the information	Provided for each exchange (as foreseen for GLAD), see the example above, for one of the exchanges: `reliability="1"` Apart from exchanges not provided for any other field

IV Life cycle and model

Time, geography and technology are provided as goal for the dataset, in the metadata, and also for each exchange, as representation, and finally also as conformance assessment, as pedigree score. But indirectly, as goal for these is mentioned in metadata of the dataset, and for each exchange, the conformance is provided as pedigree (conformance) result.

Time, goal:

```
<timePeriod startDate="2001-01-01" endDate="2006-12-31"
isDataValidForEntirePeriod="true" />
```

Location / geography, goal:

```
<geography geographyId="34dbbff8-88ce-11de-ad60-0019e336be3a">
        <shortname xml:lang="en">GLO</shortname>
        <comment>
          <text xml:lang="en" index="1">The data is based on information by a leading
international computer manufacturer. Such a laptop computer may be assembled anywhere
in the world. Therefore a global dataset is justifiable.</text>
        </comment>
```

Technology, goal:

```
  <text xml:lang="en" index="1">This dataset represents the production of 1 unit of a laptop computer. It is based on a typical  laptop computer in the last 3 years before the reference year 2005 (Pentium 3, processor speed 600 MHz, 10 GB RAM, 128 MB memory, 12.1 inch screen, total weight with expansion base  3.15 kg; including the expansion base without speaker, switch and cables). Main data are based on literature data representing a typical laptop computer of a leading producer (HP Omnibook, Hewlett Packard (HP)) and a EPD sheet. Laptop parts like hard disk drive, CD Rom drive, printed wiring boards (e.g. motherboard) and batteries are inventoried in individual ecoinvent datasets. </text>
```

Representation, time, geography, technology:

Here for one of the exchanges, polystyrene high impact (where the flow name is representing the technology)

```
<intermediateExchange id="d5e09de7-4856-40cc-9e9a-00b9663d2f97" unitId="487df68b-4994-4027-8fdc-a4dc298257b7" casNumber="9003-53-6" amount="0.42395" sourceId="c134c46b-9171-4b5f-abf3-ca7602e3c2e9" sourceYear="2003" sourceFirstAuthor="von Geibler J." intermediateExchangeId="14a9fc00-5a28-40f3-9cbe-527a40212577">
        <name xml:lang="en">polystyrene, high impact</name>
        <unitName xml:lang="en">kg</unitName>
        <comment>EcoSpold01Location=RER</comment>
        <comment xml:lang="en">Literature Value.</comment>
        <uncertainty>
          <lognormal meanValue="0.42395" mu="-0.86" variance="0.02" varianceWithPedigreeUncertainty="0.0301" />
          <pedigreeMatrix reliability="1" completeness="4" temporalCorrelation="1" geographicalCorrelation="3" furtherTechnologyCorrelation="3" />
        </uncertainty>
```

a	Time	Time or time period relevant	Provided for each exchange (as foreseen for GLAD), `sourceYear="2003"`
b	Geography	Geographical area or point relevant	Provided for each exchange (as foreseen for GLAD), `EcoSpold01Location=RER`
c	Technology	Technology or technology mix relevant	Provided for each exchange (as foreseen for GLAD), `polystyrene, high impact`

Nomenclature systems are available for the classification and for the locations, not – in the dataset – for the elementary flow exchanges; for intermediate flows, there is a classification mentioned, cpc.

```
<classification classificationId="f4a849f3-9172-404c-9834-37fae8cda5ff">
        <classificationSystem xml:lang="en">ISIC rev.4 ecoinvent</classificationSystem>
        <classificationValue xml:lang="en">2620:Manufacture of computers and peripheral equipment</classificationValue>
      </classification>
      <classification classificationId="048fe914-e1f2-46d4-a061-f37b0300f66b">
        <classificationSystem xml:lang="en">EcoSpold01Categories</classificationSystem>
        <classificationValue xml:lang="en">electronics/devices</classificationValue>
      </classification>
```

```
<comment>EcoSpold01Location=RER</comment>
```

d	Supported LCA nomenclature system(s)		Not for elementary flow based exchanges in the dataset but for intermediate flow exchanges, for the process and for the locations
e	Supported LCIA methods	See name	This is not provided in the dataset

f	Materiality, completeness	Order of 5 main drivers for main LCIA results	This is not provided in the dataset. It could be calculated though.
g	Representativeness	For science-based sampling, variation coefficient plus documentation; for expert judgement, representativeness classes estimates	Representativeness is provided as conformance assessment (see above, sample conformance), not as value; this is fine since the descriptor is static.
h	LCI modeling approach	Attributional or consequential modeling	Available for the dataset

```
<includedActivitiesStart xml:lang="en">From reception of raw materials and auxiliares at the factory gate.</includedActivitiesStart>
        <includedActivitiesEnd xml:lang="en">This dataset includes materials (mainly metals and plastics) with their respective manufacturing processes (e.g. sheet rolling, press moulding). Further inventoried is the infrastructure (factory), the electricity for the assembly of the laptop computer, the water consumption and industrial waste water, the packaging, plus the disposal of the laptop. </includedActivitiesEnd>
```

i	System boundaries	Set of criteria specifying which unit processes are part of a product system (ISO, 2006)	Criteria are not provided but the system boundaries are described
j	Method used to deal with multifunctional processes	Method used to assign the environmental burdens to the joint production of the reference flows.	Not tested since the dataset is a monofunctional process.

```
<name xml:lang="en">corrugated board box</name>
        <unitName xml:lang="en">kg</unitName>
        <comment>EcoSpold01Location=RER</comment>
        <comment xml:lang="en">Literature Value.</comment>
        <uncertainty>
          <lognormal meanValue="0.837" mu="-0.18" variance="0.02"
varianceWithPedigreeUncertainty="0.0301" />
          <pedigreeMatrix reliability="1" completeness="4" temporalCorrelation="1"
geographicalCorrelation="3" furtherTechnologyCorrelation="3" />
        </uncertainty>
        <property propertyId="6393c14b-db78-445d-a47b-c0cb866a1b25"
amount="0.433862433862434" unitId="577e242a-461f-44a7-922c-d8e1c3d2bf45">
          <name xml:lang="en">carbon content, non-fossil</name>
          <unitName xml:lang="en">dimensionless</unitName>
          <comment xml:lang="en">Biogenic C content was calculated after biogenic C
contents of chemicals used in the production of  board (Tab 10.15 &amp; 10.16,
ecoinvent v2.2 report 11_III, part 10).</comment>
        </property>
```

k	Biogenic carbon	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).	Biogenic carbon is provided as exchange property and also as separate carbon flows (which do not occur for the tested dataset)
l	Land use	See detailed format description, chapter 4	No land use for the tested dataset but the methodological background report (Weidema et al

			2013) contains a full chapter on modeling land use, pp. 47

```
<intermediateExchange id="ce5bdc37-f0cc-420a-9d9c-773cf7fcde30" unitId="487df68b-4994-4027-8fdc-a4dc298257b7" casNumber="9002-88-4" amount="0.051" sourceId="c134c46b-9171-4b5f-abf3-ca7602e3c2e9" sourceYear="2003" sourceFirstAuthor="von Geibler J." intermediateExchangeId="7ee0ddc3-924d-4339-b9bf-7856af6a698f"
[…]
        <classification classificationId="555d4947-3964-4529-ab70-20468dfee0bd">
          <classificationSystem xml:lang="en">CPC</classificationSystem>
          <classificationValue xml:lang="en">39270: Waste, parings and scrap of plastics</classificationValue>
        </classification>
        <classification classificationId="ee3238ec-ca7a-4ddc-af1b-e7c35957b9b6">
          <classificationSystem xml:lang="en">By-product classification</classificationSystem>
          <classificationValue xml:lang="en">Waste</classificationValue>
        </classification>
        <outputGroup>2</outputGroup>
      </intermediateExchange>
```

m	Wastes and end-of-life	Substances or objects which the holder intends or is required to dispose of (ISO 14040, 2006) .	Provided with a classification, albeit for the exchange not for the entire dataset. Waste modeling is clear from the ecoinvent system model.
n	Water	Water use	Water modeling not described for the dataset and also not in the background report

```
<includedActivitiesEnd xml:lang="en">This dataset includes materials (mainly metals and plastics) with their respective manufacturing processes (e.g. sheet rolling, press moulding). Further inventoried is the infrastructure (factory), the electricity for the assembly of the laptop computer, the water consumption and industrial waste water, the packaging, plus the disposal of the laptop. </includedActivitiesEnd>
```

o	Infrastructure/capital goods	Product not intended for consumption, with a lifetime exceeding one year (Weidema et al., 2013).	Documented in the dataset (but in a free text field)
p	Long-term emissions (beyond 100 years)	Emissions that will occur in the future but are determined today (EC-JRC, 2010).	Two specific categories for long term emissions, for elementary flows: output to air, low population density, long-term; output to water, ground-, long-term; report p. 65: "*Outlook:* The issue of how best to include long-term emissions is currently under consideration."
q	Temporary carbon storage, delayed greenhouse gas emissions, delayed credits for solving multifunctionality	See detailed format description, chapter 4	Not covered.

V Verification and quality assurance

The dataset is reviewed, the reviewing person is mentioned in the dataset.

```
<review reviewerId="89ae408f-f095-4276-96ca-c871888cf632" reviewerName="Roland Hischier" reviewerEmail="roland.hischier@empa.ch" reviewDate="2011-09-22"
```

```
reviewedMajorRelease="3" reviewedMinorRelease="0" reviewedMajorRevision="23"
reviewedMinorRevision="8">
        <details>
          <text xml:lang="en" index="0">
          </text>
        </details>
        <otherDetails xml:lang="en">
        </otherDetails>
      </review>
```

a	Dataset review performed	See name	Yes, available in the dataset
b	Type of dataset review	Number of reviewers and their relation to the data provider, internal or external	Reviewers available, relation to the data provider not directly but can be investigated
c	Quality assurance performed	Independent quality assurance related to the information provided, on single entries of the dataset	Pedigree quality assessment available for each exchange Not for other entries in the dataset
d	Reviewing person(s)	Person who conducted the review	Reviewers available

VI Calculation			
a	Aggregation type if any	For an aggregated dataset, specify how the aggregation was performed.	The test dataset is a unit process therefore the criterion is not applicable. For other datasets, the information is available.

VII Administrative

The dataset is not copyright-protected, therefore information about copyright-holders and so forth do not apply.

```
  <administrativeInformation>
      <dataEntryBy personId="20d13b40-bcc4-45ce-aff6-d64823ac880b"
isActiveAuthor="true" personName="Martin Lehmann" personEmail="empa@ecoinvent.org" />
      <dataGeneratorAndPublication personId="20d13b40-bcc4-45ce-aff6-d64823ac880b"
personName="Martin Lehmann" personEmail="empa@ecoinvent.org" dataPublishedIn="0"
publishedSourceId="f6bc0bb2-ee4a-4847-bc92-242dca5f5a70" publishedSourceYear="2007"
publishedSourceFirstAuthor="Hischier R." isCopyrightProtected="false"
accessRestrictedTo="1" />
      <fileAttributes majorRelease="3" minorRelease="0" majorRevision="27"
minorRevision="0" internalSchemaVersion="2.0.10" defaultLanguage="en"
creationTimestamp="2010-07-28T18:44:13" lastEditTimestamp="2013-08-09T09:59:18"
fileGenerator="EcoEditor 3.3.65.10748" fileTimestamp="2013-08-09T09:59:18"
contextId="de659012-50c4-4e96-b54a-fc781bf987ab">
        <contextName xml:lang="en">ecoinvent</contextName>
      </fileAttributes>
    </administrativeInformation>
```

a	Copyright protected	See name	isCopyrightProtected="false"

	dataset?		
b	Copyright holder	See name	n.a.
c	Free dataset or for purchase?	See name	n.a.
d	Dataset license	See name	n.a.
e	Dataset contact	See name	personName="Martin Lehmann" personEmail="empa@ecoinvent.org"

Conclusions

Overall, the test dataset contains information for many of the descriptors; the dataset is somewhat special since it is not copyright-protected (due to its age, probably).

The following descriptors are not well supported by the dataset:

1. Ig, IVe: supported LCIA methods
2. IId, IVf: model completeness conformance and materiality
3. IVn, IVq: water modeling and temporary carbon modeling

ecoinvent does not provide an explicit completeness goal, nor LCIA methods to be supported; materiality and the approach for dealing with water are rather new, and it is therefore not such a surprise that they are not well reflected in the dataset.

Multifunctionality and copyright protection could not be checked for the dataset. Long-term emissions are somehow covered but with a rather rough level of detail. Finally, some aspects fit overall but lack some specific details for a smooth recognition of the dataset; for example, reviewers relation to the data provider can be investigated based on the information provided in the dataset but is not directly stated in the dataset.

Overall, the test shows quite a broad support of the proposed descriptors, for the selected dataset. However, format issues were not dealt with, and often, the information is not directly available but needs to be extracted from text fields or from separate reports, where it is, on the other side, applicable not only for the selected dataset but for all or many datasets of the database.

5.1.2. GLAD user mock test

In a GLAD user mock test, we will assume a GLAD user and perform a fitness for purpose assessment with the test dataset, based on the proposed descriptors.

Assumed GLAD user goal and scope

Macbook Pro 2015, with the following configuration[6]:

Display size/resolution	13.3-inch 2,560x1,600 screen
PC CPU	2.7GHz Intel Core i5-5257U
PC memory	8GB DDR3 SDRAM 1,866MHz
Graphics	1,536MB Intel HD Iris Graphics 6100
Storage	128GB SSD
Optical drive	None
Networking	802.11ac wireless, Bluetooth 4.0
Operating system	Apple OS X Yosemite 10.10.2

[6] Taken from https://www.cnet.com/products/apple-macbook-pro-13-inch-2015/

The Macbook should be used in Germany, in an office for a consultancy. An impact assessment using the ReCiPe Hierarchist midpoint method should be performed. The assessment should address an average notebook of the analysed type, no market changes are anticipated, thus an attributional life cycle assessment is envisaged. No further specific requirements regarding the life cycle modeling are prescribed, but the model should be "good" and representative and follow ISO 14040.

GLAD user fitness for purpose assessment

With the rather rough goal and scope specification described in the previous section, a similarly rough, initial fitness for purpose assessment can be performed, which is shown in Table 2. Blue color in the table indicates goal and scope specification, green, yellow and red a conformance assessment.

Table 2: Initial, rough GLAD fitness for purpose assessment, Macbook

#	Metadata descriptor name	Scope	Type*	Data created by**	Descriptor segment	Mandatory?	GLAD test ffp assessment
	0 Dataset information						
a	Process name	Data set	Ad	DaP	Value	Yes	
b	Process type	Data set	Ad	DaP	Value	Yes	
	I Goal and scope						
a	Reference time	Data set	FI	GLAD	Goal	No	2016
b	Reference geography	Data set	FI	GLAD	Goal	No	German market
c	Reference technology	Data set	FI	GLAD	Goal	No	Macbook Pro 2015
d	Reference model completeness	Data set	FI	DaP	Goal	Yes	
e	Reference sample completeness	Data set	FI	DaP	Goal	Yes	
f	Intended sample approach	Data set	FI	DaP	Goal	Yes	
g	LCIA methods	Data set	FI	DaP	Goal	Yes	ReCiPe H midpoint
	II Conformance						
a	Time related conformance	Exchange	FI	GLAD	Conformance	No	11 years difference
b	Geographical conformance	Exchange	FI	GLAD	Conformance	No	data set geography includes target geography
c	Technological conformance	Exchange	FI	GLAD	Conformance	No	notebook, but distinctly different
d	Model completeness conformance	Data set	FI	DaP	Conformance	Yes	
e	Sample conformance	Exchange	FI	DaP	Conformance	Yes	
	III Reliability of the sources used						
a	Reliability of the sources	Exchange	St	DaP	Cross-cutting	Yes	
	IV Life cycle and model						
a	Time	Exchange	FI	DaP	Value	Yes	2005
b	Geography	Exchange	FI	DaP	Value	Yes	Global
c	Technology	Exchange	FI	DaP	Value	Yes	HP Omnibook
d	Supported LCA nomenclature system(s)	Data set	FI	DaP	Value	Yes	(ecoinvent)
e	Supported LCIA methods	Data set	FI	DaP	Value	Yes	n.a.
f	Materiality, completeness	Data set	FI	DaP	Value	Yes	n.a.
g	Representativeness	Exchange	St	DaP	Value	Yes	
h	LCI modeling approach	Data set	FI	DaP	Value	Yes	attributional
i	System boundaries	Data set	FI	DaP	Value	Yes	
j	Methods for multifunctional processes	Data set	FI	DaP	Value	Yes	

Conclusion

The application mock test seems useful, but shows also that many of the detailed life cycle based values / representations are not reflected in specific descriptors for goal and scope, and can therefore not be addressed directly in this, initial and rough, fitness for purpose test.

It is expected that this can to some extent be addressed in future specification and implementation of the format for the various descriptors.

5.2. Case 2: Passenger car, average

The second dataset to be investigated is an average passenger car, retrieved from the ILCD network, as PEF compliant dataset[7]. Process dataset UUID:

88f28103-2983-445b-8ff2-2033e72b8ba3.

[7] Use of this dataset in the present text also outside of the PEF context kindly granted by thinkstep AG, Dec 12 2016

5.2.1. Test for GLAD descriptor support

Overview

For the test, the original dataset in ILCD format is compared against the proposed metadata descriptors; the ILCD XML structure is shown with a

```
Brown background
```

The descriptors are shown again in a table, the same color code is used as for the previous case 1 (see section 5.1). Again, the test goes through all the descriptors proposed, as listed in Table 1. Only those descriptors are addressed which are to be provided by the data provider. As far as relevant, additional information in the dataset is commented. Focus is to investigate whether the information is available, rather than the format of the information.

Results for the descriptors

0 Dataset information

Dataset information can be easily found in the dataset, as expected.

```
<name>
        <baseName xml:lang="en">Passenger car, average</baseName>
        <treatmentStandardsRoutes xml:lang="en">technology mix, gasoline and diesel driven, Euro 3-5, passenger car</treatmentStandardsRoutes>
        <mixAndLocationTypes xml:lang="en">consumption mix, to consumer</mixAndLocationTypes>
        <functionalUnitFlowProperties xml:lang="en">engine size from 1,4l up to &gt;2l</functionalUnitFlowProperties>
</name>
[…]
<typeOfDataSet>Unit process, black box</typeOfDataSet>
```

a	Process name	Passenger car, average
b	Process type	Unit process, black box

I Goal and scope

Goal and scope for the dataset are available for some aspects; other aspects can be concluded since the dataset claims to be PEF compliant.

```
<dataCutOffAndCompletenessPrinciples xml:lang="en">A cut-off rule of 95%, based on material or energy flow or the level of environmental significance, is (if applied) clearly documented and confirmed by the reviewer, in particular with reference to the environmental significance of the cut-off applied. A cut-off rule lower than 95% is not used.
Capital goods (including infrastructures) and their End of life: they are included unless the exclusion is clearly documented and allowed according to the cut-off principles.
System boundaries: system boundaries include all known processes linked to the product supply chain.
</dataCutOffAndCompletenessPrinciples>
<deviationsFromCutOffAndCompletenessPrinciples xml:lang="en">Any capital goods such as production, end-of-life or production factories of the transport systems are not included in this dataset. In order to test the siginificance of the transport system in the specific boundaries of your study, please refer to the separate production datasets of the transport systems.
</deviationsFromCutOffAndCompletenessPrinciples>
```

```
  <common:reviewDetails xml:lang="en">[…] An active cut-off, i.e. by excluding
available data, has not been found. […]</common:reviewDetails>
```

```
<referenceToLCAMethodDetails refObjectId="f2b512cd-43b2-4260-9882-eebc06731274"
version="29.00.000" type="source data set" uri="../sources/f2b512cd-43b2-4260-9882-
eebc06731274.xml">
<common:shortDescription xml:lang="en">GaBi Modelling Principles
</common:shortDescription>
</referenceToLCAMethodDetails>
```

3.3.4 Cut-offs

Cut-off rules are defined to provide practical guidelines to be able to omit specific less relevant process chain details, while creating a specific product system. The ISO 14044 : 2006 mentions three criteria used to decide which inputs are to be included: a) mass, b) energy and c) environmental significance.

There are three different types of cut-offs:

1. A known input or substance is not connected to an upstream process chain due to lack of information
2. A known inconsistency in a mass or energy balance with a known reason
3. An unknown or known inconsistency in a mass or energy balance with an unknown reason

The GaBi database has very few cut-offs of type 1). The only two reasons for cut-offs of type 1) are mathematical starting conditions (at the very beginning of the supply chain) or confidentiality reasons of competitive formulations/substances (see table in Chapter 3.3.3). Due to the magnitude of the database content and the knowhow of our engineers, most information is available or can be developed. If a substance for which no LCA data exists is needed and is not available as a dataset, the GaBi Master database uses information for a chemically/physically-related substance and creates a "precautionary principle" scenario (rather slightly over estimate than underestimating the impact) for the substance causing the gap. If the influence of the "precautionary principle" scenario on the overall result is smaller than 5%, the scenario can stay (gap-closing insignificantly overestimates to the actual value). If the influence on the result is higher, more information is gathered or the sensitivity is quantified.

Figure 10: GaBi modeling principles (GaBi 2014, p 36)

d	Reference model completeness	Targeted coverage of inputs and outputs in the dataset	Quite complicated documentation: a 95% cut-off rule based on several possible indicators is, if used, documented (dataCutOffAndCompletenessPrinciples). No documentation could be found in the dataset. Also the in-dataset review states that a cut-off "has not been found" (reviewDetails) Therefore it seems that no cut-off has been applied, although the GaBi modeling principles, which are one of the referenced "LCA methods" referenceToLCAMethodDetails state that cut-offs may be applied in the GaBi database (Figure 10). For capital goods, it is stated that Capital goods (including infrastructures) and their End of life: they are included unless the exclusion is clearly documented and allowed (dataCutOffAndCompletenessPrinciples) but in deviationsFromCutOffAndCompletenessPrinciples this is abrogated again: Any capital goods such as production, end-of-life or production factories of the transport systems are not included in this dataset.

2.6.3. Representativeness of a PEFCR

A PEFCR is considered to be representative of a specific product category when all the following conditions are met:

1) The Technical Secretariat in charge of a specific product category has invited to contribute to the PEFCR development process all the major competitors, or their representatives (i.e. via industry associations) covering for at least 75% of the EU market (in terms of yearly turnover or production). All companies contributing to more than 10% to the EU market (in terms of yearly turnover or production) have been invited.

2) The industry stakeholders (producers/importers, either as single companies and/or as business associations) participating to the whole process cover at least 51% of the EU market (in terms of yearly turnover or production). The participation of stakeholders will be judged on the basis of their inputs to the process and/or participation to meetings. The 51% target has to be achieved by the end of the pilot phase. This means that it is not a requirement for the Technical Secretariats themselves to fulfil.

3) The Technical secretariat has invited and involved in the PEFCR development process a wide range of stakeholders, with particular reference to SMEs, consumers' and environmental associations.

Figure 11: Representativeness requirements for PEFCRs (EC 2016, pp 23)

e	Reference sample representativeness	Targeted sample representativeness, in line with the foreseen sampling approach	Not explicitly provided in the dataset; in the Product Environmental Footprint Guidance indicators for a representativeness for the European market are described (Figure 11); the same holds probably for the dataset (i.e., European market, yearly average)
f	Intended sample approach	Scientific or expert-based sampling	Not explicitly provided in the dataset; in the Product Environmental Footprint Guidance report, the foreseen assessment of reference sample representativeness follows an expert-estimate approach (with market share percentages as indicator for representativeness), Figure 11

```
<useAdviceForDataSet xml:lang="en">
[…]
The datasets are specifically designed for […] the related LCIA methods (listed below) recommended in PEF.

[…]</useAdviceForDataSet>
[…]
      <referenceToSupportedImpactAssessmentMethods refObjectId="6371834a-c6cd-4b43-9a57-41528fb6507c" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Resource depletion, mineral, fossils and renewables, midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="7fd2917c-ab26-47e8-9601-0d85880eb934" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Resource depletion water, midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="c439fa87-21a3-41ec-83dc-4a1b80a14894" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Photochemical ozone formation midpoint, human health (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
```

```
      <referenceToSupportedImpactAssessmentMethods refObjectId="97fd05d0-a026-43c7-b742-4867c7d82ed8" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Particulate matter/Respiratory inorganics midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="dd7ff158-dbdb-4f0a-93ac-8f5dee244246" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Ozone depletion midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="c4a9ba5a-26e5-490e-ad00-738d12603df9" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Land use midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="9daba0a9-bba8-4708-901b-feed65ce13ff" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Ionizing radiation midpoint, human health (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="b9da1573-83e9-4d0c-9110-9636ce1a9fb7" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Human toxicity midpoint, non-cancer effects (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="063938c7-d5d2-4150-b040-0bb719b7aa58" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Human toxicity midpoint, cancer effects (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="6d551dbf-676a-4f06-804d-85fa1d3ff671" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Eutrophication terrestrial midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="8d4488df-b524-40e5-a953-2c1e92a350c5" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Eutrophication marine midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="2617443c-175c-4b33-bfd1-0ce1d9cfbff3" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Eutrophication freshwater midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="e5d382ab-b010-4f90-be20-b7e9da57a328" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Ecotoxicity freshwater midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="2632d20e-b8ff-4083-a261-73f279ede42a" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Climate change midpoint, incl biogenic carbon (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <referenceToSupportedImpactAssessmentMethods refObjectId="4c79877f-c537-48a6-aa87-8c595ad3f657" version="29.00.000" type="LCIA method data set">
        <common:shortDescription xml:lang="en">Acidification midpoint (v1.09)</common:shortDescription>
      </referenceToSupportedImpactAssessmentMethods>
      <completenessElementaryFlows type="Noise" value="No statement"/>
    </completeness>
```

g	LCIA methods with version number to be supported	See name	Fully available in the dataset

II Conformance

```
<completenessProductModel>All relevant flows quantified</completenessProductModel>
```

d	Model completeness conformance	Coverage of inputs and outputs in the dataset in relation to the targeted coverage	See also Id, reference model completeness. Since model completeness goal and scope mixes the goal and scope for the dataset with what is present in the dataset, the entry for Id can be repeated here. In addition, it is stated that `All relevant flows quantified` – where of course "relevant" depends on goal and scope.

```
<common:reviewDetails xml:lang="en">[…] The completeness of the inventory is also very good, as the product chain from cradle to gate is well established and the emissions and resources that relevantly contribute to all product-specific relevant impact categories are well covered in the relevant processes also from cradle to gate. […].
</common:reviewDetails>
```

```
<percentageSupplyOrProductionCovered>95,0</percentageSupplyOrProductionCovered>
```

e	Sample conformance	Representativeness of the information provided, in relation to the sample conformance specified in goal and scope	Not provided for each exchange (as foreseen for GLAD), one statement in the review covers the completeness of the entire dataset, see above: `the emissions and resources that relevantly contribute to all product-specific relevant impact categories are well covered in the relevant processes` One statement regarding the representativeness of the information (95% covered which is quite excellent, for a global dataset), without further documentation. This number is also not addressed in the review statements.

III Reliability of the sources used

A list of sources used is provided, for the entire dataset, not for each exchange.

```
<referenceToDataSource refObjectId="b94fe38c-8f51-4b46-8a3c-76239a8fd287"
version="29.00.000" type="source data set" uri="../sources/b94fe38c-8f51-4b46-8a3c-
76239a8fd287.xml">
<common:shortDescription xml:lang="en">Handbuch Emissionsfaktoren des Straßenverkehrs,
Version 3.1, 2010</common:shortDescription>
</referenceToDataSource>
<referenceToDataSource refObjectId="d2737045-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d2737045-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">Directive 1999/96/EC</common:shortDescription>
</referenceToDataSource>
<referenceToDataSource refObjectId="d2737046-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d2737046-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">Richtlinie 2003/17/EG: Änderung der R 98/70/EG
über die Qualität von Otto- und Dieselkraftstoffen</common:shortDescription>
</referenceToDataSource>
```

```
<referenceToDataSource refObjectId="d2737047-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d2737047-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">EUROPA - Environment - Auto-Oil II Programme,
2000</common:shortDescription>
</referenceToDataSource>
<referenceToDataSource refObjectId="d273704a-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d273704a-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">Emissionen und Minderungspotential von HFKW, FKW
und SF6 in Deutschland Im Auftrag des UBA(29841256)</common:shortDescription>
</referenceToDataSource>
<referenceToDataSource refObjectId="d273704b-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d273704b-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">Handbuch Emissionsfaktoren des Straáenverkehrs,
Version 2.1, 2004</common:shortDescription>
</referenceToDataSource>
<referenceToDataSource refObjectId="d273704c-5fbd-11db-b0de-0800200c9a66"
version="29.00.000" type="source data set" uri="../sources/d273704c-5fbd-11db-b0de-
0800200c9a66.xml">
<common:shortDescription xml:lang="en">Umweltlexikon: Betankungsverlust,
2006</common:shortDescription>
</referenceToDataSource>
```

In addition, for each exchange, the type of data source is listed, which however does not indicate much about the reliability of the sources. For all exchanges, the entry is “mixed primary / secondary”, without further description of what is taken e.g. from a primary source. Also, it is debatable which of the sources listed for the dataset should be considered as primary source, since the ILCD handbook states that primary data sources are producers of goods and operators of processes, as well as producer and operator associations (Figure 12), which does not seem to apply for any of the sources listed.

quality as well as quality assurance of data (i.e. review) are important requirements that support valid studies. As already addressed in chapters 6.9.3 and 6.9.4, a wide range of potential LCI data sources exist:

- Primary data sources are the producers of goods and operators of processes and services, as well as their associations.
- Secondary data sources which either give access to primary data (possibly after re-modelling / changing the data) and to generic data are e.g. national databases, consultants, and research groups.

Figure 12: Definition of primary and secondary source, ILCD 2010 p 187

```
<exchange […]
        <common:shortDescription xml:lang="en">Diesel_BR</common:shortDescription>
      […]
      <dataSourceType>Mixed primary / secondary</dataSourceType>
      <dataDerivationTypeStatus>Unknown derivation</dataDerivationTypeStatus>
       […]
    </exchange>
```

a	Reliability of the sources	Reliability of the information provided in the dataset, assessed by reliability of the sources used for obtaining the information	Unclear from the dataset.

IV Life cycle and model

Time, geography and technology are provided for the dataset, in the metadata. Different from the previous dataset, this dataset does not provide goal, representation, and conformance, but instead provides "only" the representation (value) and an assessment of the overall quality of the dataset in the PEF data quality schema; it seems as if goal and representation are seen as identical for the dataset.

Time:

```
<common:referenceYear>2015</common:referenceYear>
common:dataSetValidUntil>2020</common:dataSetValidUntil>
<common:timeRepresentativenessDescription xml:lang="en">Annual average. The DQR of the dataset reflects the quality of the data at the time of release. The user of the dataset should revise the DQR at the moment of application.</common:timeRepresentativenessDescription>
```

Location / geography:

```
<geography>
      <locationOfOperationSupplyOrProduction location="GLO">
        <descriptionOfRestrictions xml:lang="en">The data set is representative for the global average.The geographical representativeness concerning the related country as well as its DQI is related to relevance of the contributors. Possible geographical variance of minor contributors below the cut-off criteria are not affecting the geographical representativeness of the overall process, but are positively affecting the completeness of the model.</descriptionOfRestrictions>
      </locationOfOperationSupplyOrProduction>
</geography>
```

The sentence "`Possible geographical variance of minor contributors below the cut-off criteria are not affecting the geographical representativeness of the overall process, but are positively affecting the completeness`" is not clear – it seems no cut-off criteria have been applied, see above, and also apart from that the meaning is not clear.

Technology:

```
<technology>
      <technologyDescriptionAndIncludedProcesses xml:lang="en">This dataset represents an average of different passenger cars. It includes both petrol and diesel cars, Euro 3 to Euro 5 cars and engine classes (displacement) ranging from 1,4l to more than 2l. This transport dataset refers to 1 km and includes the fuel. The shares have been calculated by using the gross domestic product:

Australia   3%
Brazil  4%
China   16%
Europe  30%
India   3%
Japan   8%
Russia  4%
United States 32%

Variable parameters are: sulphur content of fuel, driving share urban/interurban/motorway and total mileage over lifetime. Inputs: fuel, vehicle (production). Outputs: driving distance, combustion emissions (ammonia, benzene, carbon dioxide, carbon monoxide, methane, nitrogen oxides, nitrous oxide, NMVOC, particulate PM 2.5, sulphur dioxide, toluene, xylene), vehicle (end-of-life).
For more information please refer to: http://www.gabi-software.com/index.php?id=8375</technologyDescriptionAndIncludedProcesses>
```

a	Time	Time or time period	2015-2020, annual average

		relevant	
b	Geography	Geographical area or point relevant	Global, as weighted average of 8 regions
c	Technology	Technology or technology mix relevant	Person car, 1.4l- >2l

Nomenclature systems are not explicitly addressed in the process dataset; this is a recommended but not mandatory field in the ILCD data format and was skipped for the investigated dataset.

Field name	Element/attribute name	Requ.	Occ.	Data type
Classification information	*classificationInformation*	r	[0,1]	
Classification	*classification*	r	[0,n]	
Classification system name	@name	r		string
Classes	@classes	r		anyURI
Class name	*class*	r	[1,n]	
Hierarchy level	@level	r		LevelType
Unique class identifier	@classId	r		string

Figure 13: Classification information and related fields in ILCD 1.1

The mandatory field “Class name” exists, but renamed to common: class.

```
<classificationInformation>
        <common:classification>
          <common:class level="0" classId="7cc5e0d5-a00f-4e15-86a0-
7cd40fe932d9">Transport services</common:class>
          <common:class level="1" classId="8cc5018e-e632-4f87-915e-
ed2a6de8770c">Road</common:class>
        </common:classification>
</classificationInformation>
```

For each exchange for example, the classification system is mentioned (“GaBiCategories.xml”)

```
<dataSetInformation>
      <common:UUID>ae4ea4d8-16e3-481b-99df-3ae1c2826d96</common:UUID>
      <name>
        <baseName xml:lang="en">Diesel_CN</baseName>
      </name>
      <classificationInformation>
        <common:elementaryFlowCategorization categories="../GaBiCategories.xml">
          <common:category level="0" catId="9b374dd3-8ba9-4cf9-85aa-
75565977ae43">Energy carriers and technologies</common:category>
          <common:category level="1" catId="1400a17f-2b38-46f9-8a45-01ab383cb241">Crude
oil based fuels</common:category>
        </common:elementaryFlowCategorization>
```

d	Supported LCA nomenclature system(s)		Not at a central place in the database but available in the various datasets; interestingly it seems not at the process dataset
e	Supported LCIA methods	See name	See also Ig; the dataset often does not distinguish between goal and scope and value/representation; the LCIA methods which are stated to be supported by the dataset (see Id) are even more a statement
f	Materiality, completeness	Order of 5 main drivers for main	This is not provided in the dataset, see also IIe

		LCIA results	
g	Representativeness	For science-based sampling, variation coefficient plus documentation; for expert judgement, representativeness classes estimates	The dataset states to cover `<percentageSupplyOrProductionCovered>95,0</percentageSupplyOrProductionCovered>` (see also above)
h	LCI modeling approach	Attributional or consequential modeling	Available for the dataset `<LCIMethodPrinciple>Attributional</LCIMethodPrinciple>`
i	System boundaries	Set of criteria specifying which unit processes are part of a product system (ISO, 2006)	Quite complicated discussion in the dataset about cut-off criteria; it seems no cut-off has been applied but this is not entirely clear, see also Id
j	Method used to deal with multifunctional processes	Method used to assign the environmental burdens to the joint production of the reference flows.	Not tested since the dataset is a monofunctional process.


```
<modellingConstants xml:lang="en">Direct land use change: GHG emissions from direct LUC
allocated to good/service for 20 years after the LUC occurs.
Carbon  storage  and  delayed  emissions:  credits  associated  with  temporary
(carbon) storage or delayed emissions are not considered in the calculation of the EF
for the default impact categories.
Emissions off-setting: not included
Fossil and biogenic carbon emissions and removals: removals and emissions are modelled
as follows: All  GHG  emissions  from  fossil  fuels  (including  peat  and  limestone)
are modelled consistently with the ILCD list of elementary flows. In the case that the
emissions refer to the molecules CO2 and CH4, they are modelled as carbon dioxide
(fossil) and methane (fossil). Biogenic   uptake   and   emissions   are modelled
separately. For land use change, all carbon emissions and uptakes are inventoried
separately for each of the elementary flows. Soil carbon accumulation (uptake) via
improved agricultural management is excluded from the model.</modellingConstants>

<referenceToLCAMethodDetails refObjectId="57a2608f-18fc-454f-998a-1657376a4322"
version="29.00.000" type="source data set" uri="../sources/57a2608f-18fc-454f-998a-
1657376a4322.xml">
        <common:shortDescription xml:lang="en">GaBi Land Use Change Model
Documentation</common:shortDescription>
      </referenceToLCAMethodDetails>
```


k	Biogenic carbon	Carbon derived from biogenic (plant or animal) sources excluding fossil carbon (IPCC, 2006).	Modeling is described, as modeling constant: `Biogenic uptake and emissions are modelled separately. For land use change, all carbon emissions and uptakes are inventoried separately for each of the elementary flows. Soil carbon accumulation (uptake) via improved agricultural management is excluded from the model.`
l	Land use	See detailed format	See also above, not too clear if land use is

		description, chapter 4	considered in the dataset, since none of the exchanges in the dataset deals with land use change or occupation, but a Land Use Change Model is listed as reference.

```
< <referenceToLCAMethodDetails refObjectId="2f4e6329-f31d-44dc-8933-f4e6608590f5"
version="29.00.000" type="source data set" uri="../sources/2f4e6329-f31d-44dc-8933-
f4e6608590f5.xml">
        <common:shortDescription xml:lang="en">GaBi Water Modelling
Principles</common:shortDescription>
      </referenceToLCAMethodDetails>
```

m	Wastes and end-of-life	Substances or objects which the holder intends or is required to dispose of (ISO 14040, 2006) .	The dataset does not consider waste, at least none of the input- or output flows can be considered as waste; however, the review statement and the data quality assessment address waste and end of life modeling
n	Water	Water use	The dataset does not contain water flows. Yet still, GaBi Water Modeling Principles are referenced
o	Infrastructure/capital goods	Product not intended for consumption, with a lifetime exceeding one year (Weidema et al., 2013).	See the discussion above about cut-off criteria, Id
p	Long-term emissions (beyond 100 years)	Emissions that will occur in the future but are determined today (EC-JRC, 2010).	Not clear from the dataset

```
</deviationsFromLCIMethodApproaches>
      <modellingConstants xml:lang="en">[…]
Carbon  storage  and  delayed  emissions:  credits  associated  with  temporary
(carbon) storage or delayed emissions are not considered in the calculation of the EF
for the default impact categories.
```

q	Temporary carbon storage, delayed greenhouse gas emissions, delayed credits for solving multifunctionality	See detailed format description, chapter 4	Mentioned in modeling constants; however it is unclear whether it was applied for the dataset.

V Verification and quality assurance

The dataset states that is reviewed, by a dependent reviewer and by an independent review panel, for the panel the names are provided.

```
<validation>
      <review type="Dependent internal review">
[…]
        <dataQualityIndicators xmlns="http://lca.jrc.it/ILCD/Common">
          <dataQualityIndicator name="Technological representativeness" value="Fair"/>
          <dataQualityIndicator name="Time representativeness" value="Very good"/>
          <dataQualityIndicator name="Geographical representativeness" value="Fair"/>
          <dataQualityIndicator name="Completeness" value="Very good"/>
          <dataQualityIndicator name="Precision" value="Fair"/>
          <dataQualityIndicator name="Methodological appropriateness and consistency"
value="Fair"/>
          <dataQualityIndicator name="Overall quality" value="Good"/>
        </dataQualityIndicators>
```

```
        <common:reviewDetails xml:lang="en">[…]
In the PEF context the field "Methodlogical appropriateness and consistency" applies
solely to the criterium for implementation of the defined EoL formula, as any other
methodologial requirements are defined as mandatory.</common:reviewDetails>
[…]
      </review>
      <review type="Independent review panel">
        <dataQualityIndicators xmlns="http://lca.jrc.it/ILCD/Common">
          <dataQualityIndicator name="Technological representativeness" value="Fair"/>
          <dataQualityIndicator name="Time representativeness" value="Very good"/>
          <dataQualityIndicator name="Geographical representativeness" value="Fair"/>
          <dataQualityIndicator name="Completeness" value="Very good"/>
          <dataQualityIndicator name="Precision" value="Fair"/>
          <dataQualityIndicator name="Methodological appropriateness and consistency"
value="Fair"/>
          <dataQualityIndicator name="Overall quality" value="Good"/>
        </dataQualityIndicators>
        <common:reviewDetails xml:lang="en">[…] In the PEF context the field
"Methodological appropriateness and consistency" applies solely to the criterion for
implementation of the defined EoL formula, as any other methodological requirements are
defined as mandatory."Methodlogical appropriateness and consistency" applies solely to
the criterium for implementation of the defined EoL formula, as any other methodologial
requirements are defined as mandatory.</common:reviewDetails>
        <common:referenceToNameOfReviewerAndInstitution refObjectId="0dc85e5c-85a3-
4887-a8a6-c626f4d4661a" version="29.00.000" type="contact data set"
uri="../contacts/0dc85e5c-85a3-4887-a8a6-c626f4d4661a.xml">
          <common:shortDescription xml:lang="en">Peter Shonfield
(PhD)</common:shortDescription>
        </common:referenceToNameOfReviewerAndInstitution>
        <common:referenceToNameOfReviewerAndInstitution refObjectId="d2a7530e-3e5c-
4306-a309-41756a2c5508" version="29.00.000" type="contact data set"
uri="../contacts/d2a7530e-3e5c-4306-a309-41756a2c5508.xml">
          <common:shortDescription xml:lang="en">Johannes
Kreissig</common:shortDescription>
        </common:referenceToNameOfReviewerAndInstitution>
        <common:referenceToNameOfReviewerAndInstitution refObjectId="fc839b31-0dec-
43b0-8376-0c40a4d2ffae" version="29.00.000" type="contact data set"
uri="../contacts/fc839b31-0dec-43b0-8376-0c40a4d2ffae.xml">
          <common:shortDescription xml:lang="en">Dr.-Ing. Wolfram
Trinius</common:shortDescription>
        </common:referenceToNameOfReviewerAndInstitution>
        <common:otherReviewDetails xml:lang="en">Good overall quality (2,3) interpreted
into "good quality" in the PEF quality validation scheme</common:otherReviewDetails>
        <common:referenceToCompleteReviewReport refObjectId="7f0c8e9e-ad4c-4b78-a491-
4871caa8ad92" version="29.00.000" type="source data set" uri="../sources/7f0c8e9e-ad4c-
4b78-a491-4871caa8ad92.xml">
          <common:shortDescription xml:lang="en">GLO Passenger car, average
[agg]</common:shortDescription>
        </common:referenceToCompleteReviewReport>
      </review>
```


a	Dataset review performed	See name	Yes, available in the dataset
b	Type of dataset review	Number of reviewers and their relation to the data provider, internal or external	Reviewers available, relation to the data provider not directly but can be investigated; 2 of the 3 members of the independent review panel are employees of thinkstep, i.e. from the data provider
c	Quality assurance performed	Independent quality assurance related to the information	An overall data quality assessment is available for the dataset, it seems twice (from dependent review and from the independent panel), with identical results. Even in the

		provided, on single entries of the dataset	description, one sentence is completely identical in both reviews, with the same typos. In addition, for the panel, the same sentence without typos is provided. The assessment is not available for each exchange. By closer inspection, the panel review covers the related, aggregated dataset, and not the dataset investigated here
d	Reviewing person(s)	Person who conducted the review	Reviewers available

VI Calculation			
a	Aggregation type if any	For an aggregated dataset, specify how the aggregation was performed.	The test dataset is a unit process therefore the criterion is not applicable. For other datasets, the information is available.

VII Administrative

The dataset is copyright-protected, information about copyright-holders and so forth are available.


```
<publicationAndOwnership>
      <common:dateOfLastRevision>2016-09-01T02:00:00+01:00</common:dateOfLastRevision>
      <common:dataSetVersion>10.05.001</common:dataSetVersion>
      <common:workflowAndPublicationStatus>Data set finalised; entirely
published</common:workflowAndPublicationStatus>
      <common:referenceToOwnershipOfDataSet refObjectId="d7fa0337-f279-430d-becf-
7f3111a85010" version="29.00.000" type="contact data set" uri="../contacts/d7fa0337-
f279-430d-becf-7f3111a85010.xml">
        <common:shortDescription xml:lang="en">thinkstep</common:shortDescription>
      </common:referenceToOwnershipOfDataSet>
      <common:copyright>true</common:copyright>
      <common:referenceToEntitiesWithExclusiveAccess refObjectId="8b3e08f0-740a-43b1-
aa8d-bec56c5daf6e" version="29.00.000" type="contact data set"
uri="../contacts/8b3e08f0-740a-43b1-aa8d-bec56c5daf6e.xml">
        <common:shortDescription xml:lang="en">FINAL USER</common:shortDescription>
      </common:referenceToEntitiesWithExclusiveAccess>
      <common:licenseType>Free of charge for some user types or use
types</common:licenseType>
      <common:accessRestrictions xml:lang="en">http://www.gabi-
software.com/fileadmin/gabi/EULA_European_Commission_-
_use_of_thinkstep_LCI_data.pdf</common:accessRestrictions>
    </publicationAndOwnership>
```


a	Copyright protected dataset?	See name	Yes
b	Copyright holder	See name	available
c	Free dataset or for purchase?	See name	Depends on use type and user, but information is available
d	Dataset license	See name	Available
e	Dataset contact	See name	Available

Conclusions

Overall, also this test dataset contains information for many of the descriptors; especially, it provides the supported LCIA methods – in contrast to the previous dataset, and it provides, further, time, geography, and technology supported by the dataset, LCA nomenclature, the modeling approach (attributional), and also information about biogenic carbon, land use, and infrastructure.

On the other side, the dataset does not clearly distinguish between goal and scope, representation, and conformance assessment, for example regarding the targeted sample conformance.

For some descriptors, it is difficult to identify the information that fits for the dataset among all the information provided, see for example the discussion about the cut-off and system boundary setting applied for the dataset, above. This is amplified by often very positive statements in descriptions of the dataset. Also, several sources and modeling principles documents are listed which – probably – have not been specifically used in the context of this dataset.

More specifically, the following descriptors are not well supported by the dataset:

1. III a: reliability of the sources
2. IVf: materiality

Several of the more detailed modeling descriptors (water, long-term emissions) could not be investigated since they are probably not relevant for the dataset. Also multifunctionality could not be checked for the dataset.

Overall, this test shows that beyond the descriptor support, also the structure of the descriptors and the amount of additional information that provided in the dataset and that is not directly applicable for the dataset – is important.

As for the other dataset, format issues were not investigated during this test.

5.2.2. GLAD user mock test

Also for this dataset, we will assume the role of a GLAD user and perform a fitness for purpose assessment with the test dataset, based on the proposed descriptors.

Assumed GLAD user goal and scope

Volkswagen Golf, with the following specification[8]

Volkswagen Golf	2.0 TDI 4MOTION BMT Highline Variant
Year of production	2015
Fuel consumption	4.8 l/100km combined (5.8 city, 4.2 motorway)
Engine:	2.0 l 4-cylinder Diesel
Power:	110 kW at 3,500 r/min
Max. trailer load:	2.000 kg
Emission class:	Euro 5
Vehicle weight:	1.528 kg

The car should be used in Berlin and Germany, with an annual driving distance of 30,000 km: The analysis should cover the year 2016. An impact assessment using the ReCiPe Hierarchist midpoint method should be performed. The assessment should address an average car of the analysed type, no marginal market changes are anticipated, thus an attributional life cycle assessment is envisaged. Similar as for the other test case, no further specific requirements regarding the life cycle modeling are prescribed, but the model should be "good" and representative and follow ISO 14040.

GLAD user fitness for purpose assessment

Also for this dataset, the rather rough goal and scope specification described in the previous section

[8] Taken from a google search, https://g.co/kgs/H09skH

yields a similarly rough, initial fitness for purpose assessment, which is shown in Table 3. Blue color in the table indicates goal and scope specification, green, yellow and red a conformance assessment.

Table 3: Initial, rough GLAD fitness for purpose assessment, Golf

#	Metadata descriptor name	Scope	Type*	Data created by**	Descriptor segment	Mandatory?	GLAD test ffp assessment
	0 Dataset information						
a	Process name	Data set	Ad	DaP	Value	Yes	
b	Process type	Data set	Ad	DaP	Value	Yes	
	I Goal and scope						
a	Reference time	Data set	Fl	GLAD	Goal	No	2016
b	Reference geography	Data set	Fl	GLAD	Goal	No	German market
c	Reference technology	Data set	Fl	GLAD	Goal	No	Volkswagen Golf 7
d	Reference model completeness	Data set	Fl	DaP	Goal	Yes	
e	Reference sample completeness	Data set	Fl	DaP	Goal	Yes	
f	Intended sample approach	Data set	Fl	DaP	Goal	Yes	
g	LCIA methods	Data set	Fl	DaP	Goal	Yes	ReCiPe H midpoint
	II Conformance						
a	Time related conformance	Exchange	Fl	GLAD	Conformance	No	Same year
b	Geographical conformance	Exchange	Fl	GLAD	Conformance	No	data set geography includes target geography
c	Technological conformance	Exchange	Fl	GLAD	Conformance	No	a Golf can be considered part of a group of average cars
d	Model completeness conformance	Data set	Fl	DaP	Conformance	Yes	
e	Sample conformance	Exchange	Fl	DaP	Conformance	Yes	
	III Reliability of the sources used						
a	Reliability of the sources	Exchange	St	DaP	Cross-cutting	Yes	
	IV Life cycle and model						
a	Time	Exchange	Fl	DaP	Value	Yes	2015
b	Geography	Exchange	Fl	DaP	Value	Yes	Global
c	Technology	Exchange	Fl	DaP	Value	Yes	average car 1.4 up to > 2 l
d	Supported LCA nomenclature system(s)	Data set	Fl	DaP	Value	Yes	GaBi, ELCD
e	Supported LCIA methods	Data set	Fl	DaP	Value	Yes	The categories mentioned as being supported by the data set are partly in ReCipPe (toxicity, eutrophication); some are different (resource depletion for example).
f	Materiality, completeness	Data set	Fl	DaP	Value	Yes	n.a.
g	Representativeness	Exchange	St	DaP	Value	Yes	
h	LCI modeling approach	Data set	Fl	DaP	Value	Yes	attributional
i	System boundaries	Data set	Fl	DaP	Value	Yes	
j	Methods for multifunctional processes	Data set	Fl	DaP	Value	Yes	
k	Biogenic carbon	Data set	Fl	DaP	Value	Yes	
l	Land use	Data set	Fl	DaP	Value	Yes	
m	Wastes and end-of-life	Data set	Fl	DaP	Value	Yes	
n	Water	Data set	Fl	DaP	Value	Yes	
o	Infrastructure/capital goods	Data set	Fl	DaP	Value	Yes	
p	Long-term emissions (beyond 100 years)	Data set	Fl	DaP	Value	Yes	
q	Temporary carbon storage	Data set	Fl	DaP	Value	Yes	
	V Verification and quality assurance						
a	Data set review performed	Data set	Fl	DaP	Value	Yes	
b	Type of data set review	Data set	Fl	DaP	Value	Yes	
c	Quality assurance performed	Exchanges & data set	Fl	DaP	Value	Yes	
d	Reviewing person(s)	Data set	Fl	DaP	Value	No	
	VI Calculation						
a	Aggregation type if any	Data set	Fl	DaP	Value	If 0b) ≠ unit process	
	VII Administrative						
a	Copyright protected data set?	Data set	Ad	DaP	Value	Yes	
b	Copyright holder	Data set	Ad	DaP	Value	If VIIa) = true	
c	Free data set or for purchase?	Data set	Ad	DaP	Value	If VIIa) = true	
d	Data set license	Data set	Ad	DaP	Value	If VIIa) = true	
d	Data set contact	Data set	Ad	DaP	Value	Yes	

*ad: administrative; st: static; fl: flexible

**DaP: data provider

Conclusion

The application mock test seems again useful, here also the LCIA methods can be addressed. Sample conformance and more detailed modeling requirements were not addressed in this initial test and are left for later.

6. Metadata descriptors for enhanced searches on specific schemas

Apart from a "generic" application of LCA, where the LCA case studies are typically evaluated against ISO 14040 and 14044, there are more specific, detailed, or also focused applications which reflect specific purposes and are therefore also, in principle, calling for different or more focused fitness-for-purpose descriptors. This will be explored for two different applications in the following, for the Environmental Footprint, and for the EN 15804-compliant databases.

6.1. The Environmental Footprint

The Product and Organisational Footprint initiative of the European Commission were both initiated "with the aim of developing a harmonized environmental footprinting methodology that can accommodate a broader suite of relevant environmental performance criteria. (EC 2016a)"; meanwhile, 24 pilots for so-called representative products / representative organisations have developed models for about 80 different products / organisations, with broad industry involvement, over several years.

In this context, category rules are to be developed for each analysed product, and attributes for LCA modelling are addressed as well.

The PEF description in the Official of the European Union (EU 2013) mentions the following descriptors and elements:

- Direct "data quality criteria" (thus, fitness for purpose indicators in the context of GLAD):
    - Technological representativeness
    - Geographical representativeness
    - Time-related representativeness
    - Completeness
    - Parameter uncertainty (a different name for precision, in ISO), and
    - Methodological appropriateness and consistency

Methodological appropriateness is quite a broad criterion, and encompasses, e.g., the inclusion of capital goods (EU 2013, 5.4.2), the setting of system boundaries, and modelling of biogenic carbon (EU 2013, 5.4.9); the dataset need to be compliant with ILCD reference data and the ILCD format (Figure 14). This reflects directly the "LCA nomenclature systems" descriptor proposed here (IVd).

All relevant resource use and emissions associated with the life cycle stages included in the defined system boundaries shall be documented using the International Reference Life Cycle Data System (ILCD) nomenclature and properties ([74]), as described in Annex IV.

Figure 14: Requirement for using ILCD nomenclature and reference elements in EF studies (EC 2013, p. 33)

Overall, the detailed methodological requirements are to be developed in each pilot.

The other criteria match well with the descriptors proposed here; "completeness" fits to materiality (completeness).

In conclusion, therefore, the criteria are similar, when "methodological appropriateness" is seen as overarching criterion which contains sub-criteria (the modelling-descriptors proposed in this text). The specific definitions may be slightly different though, and will need to be aligned. This is linked to the question of a data format for the metadata which is excluded from the present text, and left for the later implementation. When aligned, the metadata descriptors of course offer the chance to directly "acknowledge" or "not-acknowledge" a dataset, without exchange of specific metadata descriptors. This brings, however, an issue of quality assurance and of trust, on the other side, since the specific descriptors and their values and assessment results are then not transparent any more.

6.2. Other specific schemes

For schemas other than EF, as discussed in the previous section, it is not expected that results will be strongly different; in principle, the descriptors proposed in this text seem supported by modern LCA data schemas; the issues lie rather in the detailed format and specification of the similarly equal descriptors and indicators, and in a mutual acknowledgement and quality assurance process which is both not available yet. Therefore, a detailed investigation of other schemas is postponed until the format for the descriptors is developed.

7. Discussion

7.1. How well are the proposed descriptors supported by ILCD and EcoSpold02

As the detailed test cases show, most of the descriptor elements are supported by the two formats considered in the analysis, ILCD and EcoSpold02. This is summarized in Figure 10.

		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		IIIa	
Calculation	Aggregation type if any		VIa	
QA	Data set review performed		Va	
	Type of data set review		Vb	
	Quality assurance performed		Vc	
	Reviewing person(s)		Vd	
Administrative	Copyright protected data set?		VIIa	
	Copyright holder		VIIb	
	Free data set or for purchase?		VIIc	
	Data set license		VIId	
	Data set contact		VIIe	

Descriptor element supported in

ILCD	ILCD & EcoSpold02	EcoSpold02	not applicable	not foreseen

Figure 15: Summary of the support by the ILCD and EcoSpold02 format for the proposed descriptors

Conformance aspects are not represented throughout both formats; LCIA methods supported are not provided in EcoSpold02, and several of the more recent modeling specifications (waste, water, biogenic carbon, long-term emissions for example) are not supported by the existing formats. Further, model completeness is not provided.

However, the results in Figure 15 are not fully representing the situation which is applicable in a real implementation, for several reasons:

- some of the elements are meant to be provided by the GLAD system and not by the data provider (1a, b, c; IIa, b, c)
- for some of the elements, the information is not available in the dataset, but in background documentation, or not for all fields in a dataset, or similar[9]; evidently, a “machine” to extract the information cannot be expected to retrieve the information from these background reports

Therefore, Figure 16 provides a more realistic overview of how well the formats considered are supporting the proposed descriptors. Partial support for a descriptor element is shown as a dotted background.

[9] these “weak deviations” have been described with red font on green background typically for the case studies

		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		IIIa	
Calculation	Aggregation type if any		VIa	
QA	Data set review performed		Va	
	Type of data set review		Vb	
	Quality assurance performed		Vc	
	Reviewing person(s)		Vd	
Administrative	Copyright protected data set?		VIIa	
	Copyright holder		VIIb	
	Free data set or for purchase?		VIIc	
	Data set license		VIId	
	Data set contact		VIIe	

Descriptor element supported in / provided by

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

Figure 16: Summary of the support by the ILCD and EcoSpold02 format for the proposed descriptors, realistic view

In this more realistic evaluation, the picture gets more colorful, but the principal conclusions remain: for the classic descriptors (called gvc descriptors in the figures), both formats show broad support, with exception of model completeness; for the more recent modeling approaches, support is less good, to lacking.

Most of the detailed modeling descriptors are part of modern methodological reports in PCRs, the Environmental Footprint initiative, or similar. Thus, for completing the descriptors, information can be taken from the methodological reports, provided it is indeed applicable for the dataset. This puts, obviously, more emphasis and responsibility in the review of the dataset.

7.2. Recommendations for completing and "populating" the descriptors

Based on the proposed descriptor short list and tests on existing datasets from two main data providers, the following points are recommended when populating the metadata descriptors:

1. **Flat descriptors**: Cross-references, from principles, to deviations from principles, to further sources, make it very difficult to understand the actual representation / value of a descriptor for a dataset. For the descriptors, a flat result must be available, which condenses the possibly various sources of input for a descriptor into the actual representation / value for the descriptor.
2. **Avoid redundant and superfluous information**: Descriptors should not contain more information than necessary; specifically, only the sources used in the dataset should be listed, and it should be clear which information has been taken from which source.
3. **Descriptor structure:** Descriptors should follow the structure of goal and scope / representation / conformance, and for each of these descriptor segments, this should be clear for the user.
4. **Evaluation statements** should only be placed in the conformance segment of a descriptor to avoid misleading claims. Descriptions should be separated from the evaluation, and be in the value and representation segment. Terms like “relevant”, “quality”, “consistent” are evaluation statements

A detailed format specification for the descriptors is left for a later implementation step. An initial proposal is provided earlier in the text. However, it is clear that the format of the descriptors should largely build on existing structures.

7.3. A short list of descriptors

All the descriptor elements proposed in Table 1 are essential for determining fitness for purpose of a dataset. However, for GLAD, the system will need to deal with datasets which are not providing all the information. The list of mandatory descriptors is very short: The name, the type of the process dataset, if applicable the aggregation type, and administrative information.

For the remaining descriptors, three requirements are to be met

1. The descriptors follow the structure proposed in this text
2. They are populated according the requirements described in section 7.2 of this text

Most of the descriptors which are not mandatory are proposed mandatory or recommended, since they are relevant for determining fitness for purpose. Those descriptors which are to be provided or specified by the user are classified as optional (from the viewpoint of the GLAD system) since it is responsibility of the user to specify what he or she needs.

Making this recommendation somewhat more prescriptive, a short list of descriptors contains those that are supported, based on the analysis in this text, by both EcoSpold02 and ILCD, which yields the descriptors listed in Figure 17. These are also listed in Table 1 as proposed mandatory.

	Goal	Value & representation	Conformance
Process name		0a	
Process type		0b	
Time	Ia	IVa	IIa
Geography	Ib	IVb	IIb
Technology	Ic	IVc	IIc
Model completeness	Id		
Sample representativeness	Ie	IVg	
LCI modeling type		IVh	
Multifunctional processes		IVj	
Biogenic carbon		IVk	
Sample approach	If		
Aggregation type if any		VIa	
Data set review performed		Va	
Reviewing person(s)		Vd	
Copyright protected data set?		VIIa	
Copyright holder		VIIb	
Free data set or for purchase?		VIIc	
Data set license		VIId	
Data set contact		VIIe	

Figure 17: Short list of descriptors

7.4. Notes on implementation and maintenance, next steps

New descriptors and a new way to deal with descriptors. The descriptor list uses several elements which are very common in LCA and data quality / fitness for purpose. These "common" elements are put in a new relation, linking goal, value/representation, and conformance. This reflects also previous work of the working group, and is essential for broadening the scope of the fitness for purpose assessment from one data base / data base network with common methodology and also format, to several interlinked data bases with potentially different methodologies and formats.

To reflect this different need further, some descriptors and elements are proposed to be applied in a way which is somewhat different from current practice, especially:

- Materiality
- Representativeness
- Quality assurance

Limitations of the shortlist. The shortlist does not contain more detailed "requirement sets" specified by established schema, such as ILCD entry level compliance, PEF compliance, or also requirements put forward by initiatives such as GHG reporting or GRI reporting. Specifying these requirement sets is expected to make the application of the descriptor list faster and easier, since several of the descriptors, e.g. for the more detailed LCI modeling questions, can be completed taking the entries for the requirement scheme. This will be elaborated in upcoming task 3.3 of the WG. Further, neither the shortlist in section 3 nor the more detailed requirements and format specification in section 4 provides detailed format specifications for the descriptors. This is left for future implementation; the WG indicated a tentative, pseudo format to hopefully support the more detailed implementation later. To this end, also cross-references to existing format implementations are provided, for several selected descriptors.

Mandatory and conditionally mandatory. The descriptor shortlist contains also an evaluation whether descriptors are seen as mandatory. The majority of descriptors is mandatory. A later implantation could develop default values for the descriptors (e.g. for the review: no review performed), which is expected to reduce the effort for providing descriptors. Several of the entries I are “conditionally mandatory”: some few descriptor entries can become mandatory depending on entries for other descriptors. For example, information regarding the aggregation is only relevant for aggregated datasets (and if a dataset is aggregated, this information is mandatory).

7.5. Topics not dealt with

7.5.1. Aggregating descriptors

An aggregation of metadata descriptors is relevant for determining the metadata descriptor representation for an aggregated dataset, when the metadata descriptors of the processes included in the aggregation are known.

How to aggregate metadata, data quality, or fitness for purpose descriptors is nowadays still an open question. Broadly established data quality schemas rely on expert judgment without providing further guidance on how the aggregation should be performed.

For this reason, the aggregation of metadata descriptors is not considered in the present text.

7.5.2. Specific schemas

The descriptors now are designed to be open to any specific modeling requirement in LCA. Recent practice in LCA has led to development of specific schemas, like EPDs, PCRs, PEF. These are not investigated in the present text. It is expected that by considering these specific schemas, several of the descriptors can be populated “at once” since they are bound to a specific schema, for example regarding specific modeling aspects of end-of life or of supported LCIA methods. However, this puts more emphasis and responsibility on the review of the datasets, and the investigated datasets suggest that the review can be further improved. Therefore, a consideration is not trivial. It is further expected, though, that such an extension to specific schemas and their “acknowledgement” can make the population and use of the datasets more efficient.

8. Glossary

Aggregated dataset. A process dataset showing the aggregated environmental exchanges and impacts of the product system related to one specific product from the activity (Weidema et al. 2013).

Aggregation. The action of summing or bringing together information (e.g., data indicator results) from smaller units into a larger unit. (e.g., from inventory indicator to subcategory) (Benoit and Mazijn 2009).

Conformance assessment. Assessment of the ability of data to fit conform to a specified goal and scope.

Data field. Container for specified data with a specified data type. (ISO/TS 14048:2002)

Data quality indicator. See definition for descriptor.

Data quality scheme. A structured system which assesses the data quality of a dataset, either quantitatively or qualitatively (e.g. pedigree matrix).

Data quality. Characteristics of data that relate to their ability to satisy stated requirements. (ISO 2006)

Data. Factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation (Merriam-Webster 2016)

Descriptor. Metadata information that describes the fitness for purpose of data. Descriptors can inform users of the model, process or flow dataset(s).

Dynamic property. A data property that changes depending on the situation in which the data is being used, or each time the data quality goals are changed. These properties should be completed for each uses of the data (e.g. representativeness).

Goal and scope. The first phase of an LCA; establishing the aim of the intended study, the functional unit, the reference flow, the product system(s) under study and the breadth and depth of the study in relation to this aim. (Guinée et al. 2002)

Inventory dataset. A process dataset which contains inventory information, in contrast to EPD datasets and LCIA datasets which contain also LCIA result information.

Life cycle database. A system intended to organize, store, and retrieve large amounts of digital LCI datasets easily. It consists of an organized collection of process datasets that completely or partially conforms to a common set of criteria, including methodology, format, review, and nomenclature, and that allows for interconnection of individual datasets that can be specified for use with identified impact assessment methods in application of life cycle assessment and life cycle impact assessments. (UNEP/SETAC 2011)

Material. Material elements of an LCA dataset are those that could make a major difference to the results of an LCA that uses the dataset, or are likely to provide the basis for conclusions drawn from such an LCA. Material elements may be within the documentation, the flow data, or the uncertainty information (Ciroth et al. 2016).

Metadata. Data that defines and describes the content of a dataset, and allows users to understand the fit for purpose (GLAD WG1 2016, key concept 'metadata'). Typically, in a process dataset all data that is not exchanges, units and amounts.

Model descriptors. (Currently outside the scope of this task and to be defined in future work).

Process dataset. A set of input and output flow data, all related to the same reference flow of a process, and containing metadata. The process can be a unit process or an aggregated process (Adapted from UNEP 2011)..

Raw data. Data used in unit process inventory modelling 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 (UNEP 2011).

Representativeness. Qualitative assessment of the degree to which the dataset reflect the true population of interest (geographic, temporal, technological, statistical). (ISO 14040:2006)

Static property. A data property that is not situationally dependent. A property of the data that never changes (e.g. reliability - because the data generation method will not change unless new data is used).

Unit process model. A group of mathematical relations that transforms raw data into a unit process dataset (UNEP 2011).

User. Person or organisation responsible to construct an LCA model from one or more unit process datasets and/or aggregated process datasets taken from databases and/or personal or organizational investigations. The user is responsible for presentation and interpretation of the LCA results and the linked recommendations within a decision process. The user is not necessarily the decision maker. (UNEP 2011)

Validation. Ensuring that data satisfy defined criteria. (UNEP 2011)

Verification. Confirmation, through the provision of objective evidence that specified requirements have been fulfilled. (UNEP 2011)

9. References

Allacker K., Souza, D.M., Salla S. (2014) Land use impact assessment in the construction sector: an analysis of LCIA models and case study application. Int J Life Cycle Assess 19:1799-1809. doi: 0.1007/s11367-014-0781-7.

Bakst JS, Lacke CJ, Weitz KA, Warren JL (1995) Guidelines for Assessing the Quality of LCI Analysis. U.S. Environmental Protection Agency, Washington, D.C.

Benoit C, Mazijn B (2009): Andrews, E. S., Barthel, L.-P., Beck, T., Benoit, C., Ciroth, A., Cucuzella, C., Gensch, C.-O., Hérbert, J., Lesage, P., Manhart, A,. Mazeau, P., Mazijn, B., Methot, A.-L., Moberg, A., Norris, G., Parent, J., Prakash, S., Reveret, J.-P., Spillemaeckers, S., Ugaya, C. M. L., Valdivia, S., Weidema, B.: UNEP/SETAC Life Cycle Initiative: Guidelines for social life cycle assessment of products, 2009

Brandão M., Levasseur A., Kirschbaum M.U.F., Weidema B.P, Cowie A.L., Jørgensen S.V., Hauschild M.Z., Pennington D.P., Chomkhamsri K. (2013) Key issues and options in accounting for carbon sequestration and temporary storage in life cycle assessment and carbon footprinting. Int J Life Cycle Assess 18:230-240. doi: 10.1007/s11367-012-0451-6.

Ciroth et al. 2008: Ciroth, A., Srocka, M.: How to Obtain a Precise and Representative Estimate for Parameters in LCA: A case study for the functional unit, in: Int J of LCA 13 LCA (3) pp 265 - 277 (2008).

Ciroth Lundie Huppes 2008: Ciroth, A.; Lundie, S.; Huppes, G.: Inventory Methods in LCA: Towards Consistency and Improvement, UNEP-SETAC Life Cycle Initiative life cycle inventory (lci) programme, task force 3: methodological consistency, VDM 2008.

Ciroth, A., Hildenbrand, J., Zamagni, A., Foster, Chr.: Life Cycle Inventory Dataset Review Criteria Development, final report v. 3.1, August 2015.

Ciroth, A., Zamagni, A., Foster, Chr., Hildenbrand, J.: Life Cycle Inventory Dataset Review Criteria Development, Review and Shonan Global Guidance Principles Criteria, gmvrp, goal, model, value, relevance, and procedure, final report v. 2.3, July 2016.

EC 2016: Product Environmental Footprint Pilot Guidance, Guidance for the implementation of the EU Product Environmental Footprint (PEF) during the Environmental Footprint (EF) pilot phase, Version 5.2, 2016, http://ec.europa.eu/environment/eussd/smgp/pdf/Guidance_products.pdf.

EC 2016a: European Commission, Environment, Singe Market for Green Products, policy background, website last update 20/06/2016 http://ec.europa.eu/environment/eussd/smgp/policy_footprint.htm.

EC-JRC (2010) General guide for Life Cycle Assessment - Detailed guidance. ILCD Handbook—International Reference Life Cycle Data System, European Union EUR24708 EN. ISBN 978-92-79-19092-6. At http://publications.jrc.ec.europa.eu/repository/handle/JRC48157.

ELCD 2016: Process dataset kaolin coarse filler, http://eplca.jrc.ec.europa.eu/ELCD3/resource/processes/4a1ebe7c-6835-4a22-8b2e-3201f1cd32e8, access July 1 2016.

ELCD 2016b: Flow dataset bauxite, http://eplca.jrc.ec.europa.eu/ELCD3/resource/flows/08a91e70-3ddc-11dd-97dd-0050c2490048.xml, access July 1 2016.

EU 2013: Official Journal of the European Union L 124/1, 5.2013, online http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013H0179&from=EN.

Frischknecht R., Althaus H.-J., Bauer C., Doka G., Heck T., Jungbluth N., Kellenberger D. and Nemecek T. (2007) The environmental relevance of capital goods in life cycle assessments of products and services. . Int J Life Cycle Assess OnlineFirst, pp., from DOI: http://dx.doi.org/10.1065/lca2007.02.309.

Frischknecht, R., Jungbluth, N.: (edts): Althaus, H.-J., Doka, G., Dones, R., Heck, Th., Hellweg, St., Hischier, R., Nemecek, Th., Rebitzer, G., Spielmann, M., Wernet, G.: Overview and Methodology Data v2.0 (2007), ecoinvent report No. 1, Dübendorf 2007.

GaBi 2014: GaBi Database & Modelling Principles 2014, PE International, November 2014, www.gabi-software.com/fileadmin/gabi/Modelling_Principles/GaBi_Modelling_Principles_2014.pdf.

GLAD WG1 2016: Network Architecture co-chairs and Christoffer Krewer: Statement of Objectives of the Global LCA Data Access (GLAD) solution, v. 1.5, May 24 2016.

GLAD WG3 2016: Bruce Vigon, Andreas Ciroth (co-chairs) with input from WG members: Meta-Data Needs Assessment – Element 1, v. 3, March 2016.

Greenberg J (2003) Metadata and the World Wide Web. In: Drake MA (ed) Encyclopedia of Library and Information Science vol 3. Marcel Dekker, Inc., New York, p 1876.

Guinée JB, Gorrèe M, Heijungs R, Huppes G, Kleijn R, Koning Ad, Oers Lv, Wegener Sleeswijk A, Suh S, Udo de Haes HA, Bruijn H, Duin Rv, Huijbregts MAJ (2002) Handbook on life cycle assessment. Operational guide to the ISO standards. Eco-Efficiency in industry and science, 7. Kluwer Academic Publishers, Dordrecht (NL).

Hansen et al. 1953: Hansen, M.H., Hurwitz, W.N., and Madow, W.G., Sample Survey Methods and Theory, Volumes I and II, John Wiley & Sons, New York, 1953.

IPCC (2006) 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.

ISO 14040 (2006) Environmental management - Life Cycle Assessment - Principles and Framework. International Standards Organization, Geneva (Switzerland).

ISO 14044 2006: Environmental management - Life Cycle Assessment - Principles and Framework. International Standards Organization, Geneva (Switzerland).

ISO/IEC (2004) 11179-1:2004 Information technology--Metadata registries (MDR)--Part 1: Framework. International Standardization Organization.

ISO/TS (2002) 14048:2002(E), Environmental management -- Life cycle assessment -- Data documentation format.

Koellner T., de Baan L., Beck T., Brandão M., Civit B., Margni M., Milà i Canals L., Saad R., Souza D.M., Müller-Wenk R. (2013) UNEP-SETAC guideline on global land use impact assessment on biodiversity and ecosystem services in LCA. Int J Life Cycle Assess 18:1188-1202. doi: 10.1007/s11367-013-0579-z.

SETAC 1994: Fava, J.; Jensen, A.A.; Lindfors, L.; Pomper, St.; de Smet, B.; Warren, J.; Vigon, B. (edts.): Life-Cycle Assessment Data Quality: A Conceptual Framework, workshop report of the workshop in Wintergreen, VA, October 1992, published by SETAC, Pensacola 1994.

UNEP 2011: Sonnemann, G., Vignon, B. (edts.) (2011): Vigon, B., Curran, M., Sonnemann, G., Wang, H., Ciroth, A., Broadbent, C., Stevenson, M., Inaba, A., de Beaufort, A., Fava, J., Draucker, L., Goedkoop, M., Baitz, M., Frischknecht, R., Krinke, S., Suppen, N., Weidema, B., Wolf, M.-A.: Global Guidance Principles For Life Cycle Assessment Databases - A Basis for Greener Processes and Products; UNEP.

US EPA (2000) Guidance for Data Quality Assessment: Practical Methods for Data Analysis. QA00 Update edn. United States Environmental Protection Agency, Washington, DC.

USDA 2016: LCI schema data dictionary v. 1.0.2, Excel workbook, 2016.

Weidema B.P., Bauer C., Hischier R., Mutel C., Nemecek T., Reinhard J., Vadenbo C.O., Wernet G. (2013) Overview and methodology. Data quality guideline for the ecoinvent database version 3. Ecoinvent Report 1(v3). St. Gallen: The ecoinvent Centre.

WG 3 2015: Meta-data Working Group Work Plan 2015-2017, version 1.20, September 9, 2015.

Yin, R.K.: Case study research, design and methods; Sage 2003.

Yodkhum, Sampattagul 2014: Yodkhum, S., Sampattagul, S.: Life Cycle Greenhouse Gas Evaluation of Rice Production in Thailand, proceedings, ENRIC 2014, November 2014.

10. Annex

10.1. ISO 14040 goal and scope explored

To further explore implications of ISO 14040 and 14044 goal and scope, Figure 18 shows the points mentioned in the interpretation section of ISO 14044 which might be relevant for a description and "fitness for purpose evaluation" of datasets. The background color refers to whether ISO sees elements as mandatory or optional (shall, shall be considered, should), which is of course not directly applicable for datasets but gives a possible direction.

ISO 14044 Life Cycle Interpretation that might be related to datasets

Legend shall | shall be considere | should

Nr	detailed in item	In Goal and Scope	Item	related to	Observation
4.5.1.1			Significant issues		which are the unit processes that contribute the most. Beware as for a dataset, impact is not taken into account.
4.5.1.1	4.5.3.1	4.2.3.6.2	Completeness check		As for the dataset, this might be related to all inputs, outputs.
4.5.1.1	4.5.3.1		Sensitivity check	significant inputs, outputs and methodological choices allocation, cut-off criteria, boundary and system, value judgement, assumptions, data quality	Shall include interpretative statements with detailed application of sensitive analysis to comparative assertions to be disclosure to the public
4.5.1.1	4.5.3.1	4.2.3.6.2	Consistency check		
	4.5.3.4	4.2.3.6.2	geographical coverage consistently applied	modelling choices (besides allocation and system boundary), data source, temporal, geographical and technological coverage	See Example consistency check
	4.5.3.4	4.2.3.6.2	temporal coverage consistently applied		
	4.5.3.4		allocation rules consistently applied		
	4.5.3.4		system boundary consistently applied		
4.5.3.1	4.5.4		Conclusions		consistent with the goal and scope
4.5.1.1	4.5.1.2		Limitations	limitations identified by data quality assessment and sensitivy analysis	
4.5.1.1	4.5.4		Recomendations		Related to the conclusions and intended application.
4.5.1.2			appropriateness of the function		
4.5.1.2			appropriateness of the functional unit		
4.5.1.2	4.5.2.3		appropriateness of the system boundary		
4.5.2.3			allocation rules		
4.5.2.3			other methodological choices		
4.5.3.1	4.5.3.3	4.2.3.6.2	uncertainty analysis		
4.5.3.1	4.5.3.4		data quality analysis	difference in data quality	
4.5.3.4			assumptions		
4.5.3.4			methods		
4.5.3.4			data		

Figure 18: ISO 14044 interpretation aspects that might be related to an evaluation of datasets

10.2. An example of assessing consistency in a dataset

An example of a Goal and Scope of a dataset is shown in Figure 19. In this example, the goal of the dataset is that it is available to be used for comparative assertion. The requirements of the scope include primary data for foreground processes, the current (2012 – 2015) technology used in companies that represent at least 70% of the Brazilian production. The dataset should use attributional modelling, either for allocation procedure (mass) or for aggregation (both partial and complete aggregation). Furthermore, the cut-off criteria used is 1% of mass or energy. The dataset must clarify what are the processes included, in the case of unit processes or partially aggregated datasets. In the case of datasets that include links to other databases, this should also be explicit.

One point explicitly mentioned in the interpretation section in ISO 14044 is consistency (4.5.1.1, see Figure 18). The aim of this text is to try out the evaluation of the consistency of a dataset according to the Goal and Scope of the dataset or a database.

	Example: Goal and Scope of a dataset		
Goal of the dataset	can be used in comparative assertion		
data source	primary data for foreground processes		
technology coverage	average of current practices of 5 companies (70% of the Brazilian production)		
time coverage	2012 – 2015		
geographical coverage	minimum of 70% of Brazilian production volume		
multifunctionality	attributional	**allocation rule**	mass
cut-off criteria	1% of mass / energy		
boundary	processes included	for partially aggregated and unit processes datasets	
	preparation of the land, foreground processes, link to background processes	for fully aggregated datasets	

Figure 19: Example of Goal and Scope of a dataset

To evaluate the consistency of a dataset according to the goal and scope, an extraction of a unit process dataset metadata is shown in Figure 20. Although there is information for each flow, the consistency is only evaluated to the dataset, shown in blue. In this example, it was only counted the % of flows that conformed to the Goal and Scope (all the 3 inputs were from primary data, however, only 2 outputs were based on primary data). How conform is it to the Goal and Scope of the Dataset? In this case, 75% of the flows are conform.

As for modelling choices, one of the outputs used emissions due to iLUC (consequential approach).

Inputs	**Data source**	**technology coverage**	**time coverage**	**geographical coverage**	**modelling choices**
A	Primary data	average of the sample (5 companies)	2015	70% of Brazilian production	
B	Primary data	best case	2015	10% of Brazilian production	
C	Primary data		2015		
Outputs					
U	Primary data	one company	2015	10% of Brazilian production	
V	Primary data	one company	2015	10% of Brazilian production	
W	Literature	experimental process	2010	50%, based on expert judgement	
X	Literature	same technology	2015, based on	Europe	average (attributional)
Y	Emission modelling		2015, based on	10% of Brazilian production	
Z (emissions due to LUC	Primary data		2015, based on	input data C	included iLUC (consequential)
CONFORMANCE TO DATASET GOALS	100% of inputs, 50% of outputs	1/3 of the inputs, 0 outputs	100% of inputs, 83% of outputs	1/3 of inputs and very poor to outputs.	1/6 of the outputs are not in conformance (iLUC)
OBSERVATION	in this example, only counted. It could be % of mass, energy... Or a qualitative assessment	A qualitative assessment may be very poor.		A qualitative assessment may be very poor.	

Figure 20: Extraction of a unit process dataset – example

However, when partial aggregation is performed, more information is needed, e.g. the flows among the processes, which is shown in Figure 21 (two unit processes, UP I and UP II). There are four possibilities related to the information regarding these flows: all the information (quantitative and qualitative is available, only the links are available, without clear information on quantities, the names of the processes are listed without the flows, and a black box dataset as fourth option). Furthermore, modelling choices could be applied differently. In UP I, flow B, system expansion was applied whereas in UP II, an attributional approach was used.

Inputs	Data source		modelling choices	
Included processes	**UP I**	**UP II**	**UP I**	**UP II**
A	Primary data	Primary data	attributional	attributional
B	Primary data	Primary data	process expansion	attributional
C	Primary data	Primary data	attributional	attributional
I		**Primary data**		
Outputs				
I	Primary data			
II		Primary data		
X	Primary data	Literature		
Y	Literature	Estimated		
Z	Primary data	Similar process		
CONFORMANCE TO DATASET GOALS	all the processes were included and there is quantitative information of the intermediary flows		there is difference on the application of multifunctionality solution for one flow	
OBSERVATION				

Figure 21: Extraction of a partially aggregated dataset - example

This could be even more relevant when full aggregation is performed. In Figure 22, the dataset is linked to only one database and Figure 23, with three databases. In both cases, the evaluation should be taken into account if the Goal and Scope of the Database is consistent to the Goal and Scope of the Dataset.

Inputs	Data source	linked to
A	Primary data	Database I
B	Primary data	Database I
C	Primary data	Database I
Outputs		
X	Primary data	
Y	Literature	
Z	Emission modelling	
CONFORMANCE TO DATASET GOALS		
OBSERVATION		Does Database I fit to the Goal of the dataset?

Figure 22: Extraction of a fully aggregated dataset linked to one database for the background processes- example

Inputs	Data source	linked to	Observation
A	Primary data	Database I	
B	Primary data	Database J	Includes infrastructure
C	Primary data	Database K	
Outputs			
X	Primary data		
Y	Literature		
Z	Emission modelling		
CONFORMANCE TO DATASET GOALS OBSERVATION		Very poor consistency What are the differences among the Databases to the Goal of the dataset?	

Figure 23: Extraction of a fully aggregated dataset linked to more than one database for the background processes - Example

The assessment of "very poor consistency" obviously depends on goal and scope of the database.

10.3. Internal version changes

v. 11.2: Comments Edivan Cherubini, Bruce Vigon included.

v. 11.1: Additional page with minimum list; edits from Peter Arbuckle included, dataset changed back to data set in ILCD original texts

v. 11: review and stakeholder feedback considered, e.g. from Bo Weidema

v. 10.3: text merge, 10.2 & commented version Bruce Vigon, plus final author group comments considered.

v. 10.2: comments Llorenç Mila I Canals considered

v. 10.1: GLAD descriptors explained, nomenclature corrected

v.10: New section metadata descriptors for specific schemas; recommended mandatory added as "necessity", executive summary extended.

v.9.1: "recommended" as status for descriptor elements introduced, minor changes, typos corrected

v.9: Executive summary added, chapter 2 shortened, descriptor definitions completed, comparison added, minor changes, typos corrected

v. 8: Concept of self-declared representation, administrative descriptors, examples added, glossary updated

v. 7.2: Changes based on inputs from Ashley Edelen and Peter Arbuckle and Andreas Ciroth

v. 7: Reviewer comments have been addressed, see also review reply document

v. 6.: Separation of sections 4 and 5, completion of several chapters including 3, 6, 7: Andreas; comments and extensions: Peter, Edivan

v. 5: contributions integrated from Cassia, Peter, Ashley, Edivan

v. 4: contributions integrated from Cassia (ISO 14040 goal and scope), Peter, Ashley, Edivan (format of descriptors)

v. 3: Comments from Cassia considered, major comment from Peter addressed with a note only (ISO, rationale); section 4 (descriptor use) added

v. 2:

- smaller additions and corrections in section 2,
- descriptors proposed in section 3, some with details
- glossary started by Ashley Edelen