GUIDANCE ON GLAD'S METADATA DESCRIPTORS

Version 1.0
Release date: 03-2023

Document developed by GLAD's Metadata Working Group

Acknowledgements

Authors

Carolina Scarinci (Chilean Network of LCA, Chile)
Peter Arbuckle (US Department of Agriculture, USA)
Pieter Callewaert (NORSUS, Norway)
Carl Vadenbo (ecoinvent, Switzerland)

Contributions and revision

Oliver Kusche (Oliver Kusche Research & Consulting, Germany)
Pier Luigi Porta (ENEA, Italy)
Tiago Emmanuel Nunes Braga (IBICT, Brazil)
Simone Fazio (ecoinvent, Switzerland)
Antonio Valente (European Commission, Joint Research Centre, Italy)

The **Life Cycle Initiative** hosted by the **UN Environment Programme (UNEP)** is the custodian of GLAD. Strategic decisions are taken by the Steering Committee of the Life Cycle Initiative, which is technically advised by the **Technical Management Group** of GLAD. The development of this guidance was originated within the activities of GLAD's **Metadata Working Group** (WG2)

Members of GLAD's Metadata Working Group

Chair: Carolina Scarinci (Chilean Network of LCA) / Peter Arbuckle (USDA)
Pieter Callewaert (NORSUS, Norway)
Carl Vadenbo (ecoinvent, Switzerland)
Erin Antognoli (USDA, USA)
Elsa Valencia Martinez (PRé Sustainability, Netherlands)
Ezra Khan (USDA, USA)
Simona Scalbi (ENEA, Italy)
Antonio Valente (European Commission, Joint Research Centre, Italy)

Members of GLAD's Technical Management Group

Claudia Giacovelli (UN Environmental Program, France)
Llorenç Mila I Canals (UN Environmental Program, France)
Jonathas De Mello (UN Environmental Program, France)
Robert Reinhardt (UN Environmental Program, France)
Carolina Scarinci (Chilean Network of LCA, Chile)
Peter Arbuckle (USDA, USA)
Pieter Callewaert (NORSUS, Norway)
Carl Vadenbo (ecoinvent, Switzerland)
Tiago Emmanuel Nunes Braga (IBICT, Brazil)
Simone Fazio (ecoinvent, Switzerland)
Koichi Shobatake (TCO2 Co. Ltd, Japan)
Antonio Valente (EC Joint Research Centre, Italy)

Content

1. Introduction

The Global LCA Data Access

Life Cycle Assessment (LCA) is a widely used methodology to assess the impact of human's activities on nature by quantifying the externalities of production and consumption. LCA has been proposed as an appropriate methodology to create specific business-relevant indicators to measure and achieve the targets set in the Sustainable Development Goals. However, mainstream uptake of LCA might be hampered by poor availability and access to data.

The Global LCA Data Access (GLAD) network aims to increase data accessibility and interoperability, offering benefits for different user profiles: users from industry can base their sustainability decisions on more diverse and accurate information; researchers can obtain more comprehensive results for their studies in a more time-efficient way; and governments can enhance public policy related to Sustainable Consumption and Production, Climate Change, Circular Economy and Environmental Labelling.

The GLAD platform launched a functional beta version in 2018 for iterative development based on real use, and was officially launched in 2020. It enables users to search, filter, compare, and access life cycle inventory datasets from different sources in a single open access digital interface. Databases continue to be operated independently by each data provider who becomes a node of GLAD. To become a node, that is to connect their database to GLAD, providers need to comply with a set of minimum metadata descriptors. GLAD provides open access to the metadata for each dataset connected, and directs users to the node's website for downloading datasets, which can be available either for free or for a licence fee. Nodes connected to GLAD benefit from increased visibility in the LCA community, with an increasing number of users visiting and downloading such data.

GLAD is accessible under http://www.globallcadataaccess.org link. Data providers wishing to publish and make their data available to users worldwide through GLAD, can find more information here.

Metadata descriptors

Life cycle inventory (LCI) datasets are models that represent commercial and industrial processes and systems. What we refer to as LCI data are rarely comprised solely of measured data, but are derived from a combination of measured data and calculations to produce estimates for material and energy flows to and from the environment. Therefore, complete documentation of the methods and assumptions used to derive results are critical for the interpretation and reproducibility of results and re-use of models. In fact, LCI models are predominantly comprised of metadata which describe the processes being modelled. 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 example, quantities of environmental exchanges are described by their reference flows which include flow name, compartment, unit of measure and a persistent identifier (if managed in software). Most of the information is contained in the reference flow, the quantity is added upon measurement or calculation. Further, metadata are not only critical for users to understand how models are built, but complete and structured metadata are critical for use in modelling software and information systems, like GLAD. Metadata elements such as process name, flow name, actors, sources, and units of measure, to name a few, are referenced with unique identifiers so that datasets can be managed and used in software.

In June 2017, a document was created by GLAD's WG3[1] to provide the core metadata descriptors and guidance on populating descriptors. This document provides the minimum set of metadata descriptors, readily applicable by the main existing LCA databases, that constitute the basic metadata descriptors required from datasets to be connected to GLAD in order to allow for interoperability. It also sets out approaches for improving assessment of fitness for purpose, providing 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.

[1] Members of WG3: Andreas Ciroth, Peter Arbuckle, Edivan Cherubini, Cassia Ugaya, Ashley Edelen

About the Guidance

The purpose of this document is to update and simplify the existing guidance about GLAD's metadata descriptors. It is meant to provide users of GLAD guidance about the metadata descriptors that are used to catalogue data within GLAD. "Users of GLAD" refers both to data providers in the process of linking datasets to GLAD and end users who look for and access data via GLAD. It is also a relevant document for software providers who aims to make their software compatible with the GLAD platform in the future.

This guidance was elaborated by WG2 of the Global Life Cycle Data Access Network[2], based on previous guidance for metadata descriptors[3]. WG2 has defined a classification structure to determine the relevance of the metadata descriptors used in GLAD.

Descriptors are classified in:

1. **Mandatory (M):** These are the minimum descriptors required for nodes wanting to connect to GLAD.
2. **Recommended (R):** These are the descriptors that are recommended by the MDWG to improve data findability as they are linked to the fields in the faceted search
3. **Optional (O):** These descriptors are meant to provide users additional information so that they can assess if a particular dataset is useful for their purpose.

Section 2 of this guidance presents the complete list of metadata descriptors used in GLAD. Section 3 and section 4 provides guidance on descriptors which are mandatory and recommended, respectively. Finally, Section 5 provides the list of optional descriptors with a brief description.

For each mandatory and recommended metadata descriptor, guidance is provided in the following order:

- **General guidance related to UI:** indicates the definition and content of the descriptor and how it is exhibited in the user interface (UI) of GLAD
- **Guidance related to API for data providers:** indicates how data providers should complete the information for the descriptor either through the GLAD API or using the Document GLAD_XLS_Template_v007.xlsx
- **Definitions:** if required, some descriptors include definitions related to LCA modelling.

[2] Members of WG2 are detailed in the Acknowledgements section
[3] Supporting document for nodes interested to join GLAD, October 2018

Updates to the Guidance

Updated to this Guidance will be performed every time modifications to the API and/or UI are undertaken, and are responsibility of the Metadata Working Group. The table below shows the records of the version history.

Version	Date	Description of the modification
0.1	12-2022	Final draft from the Metadata Working Group
1.0	03-2023	First issue revised by the TMG

How the Guidance was developed

This Guidance was developed by GLAD's Metadata Working Group (WG2) during 2022. The developed document was revised by GLAD's Technical Management Group (TMG) before the final edition.

In order to update the existing metadata descriptor guidance, the WG2 worked during 2021 in the verification of the functional requirements of GLAD, from the perspective of two types of users. On the one hand, it assessed if the end user requirements were being met by the current user interface through a series of focus groups. Furthermore, it assessed the requirements of data providers and how they were met through existing descriptors, through individual interviews.

Overall, 22 participants representing users participated in 2 online focus groups that took place between September and December 2021. Users represented mainly universities and research centers, as well as consultants, from Europe, Asia, Latin America and USA. One of the main results obtained is that GLAD possesses great value as an entry point to find data and users highly appreciate a powerful search function to ensure data findability.

Representatives from a total of five nodes were interviewed with a semi-structured instrument that aimed to collect information about the type of organisation and mission as it pertains to LCA; the motivations for participating, expectations from GLAD, and measuring success; and the nature of LCA Program: customers, users, workflows, data types, etc. The institutions that answered the interview were: ecoinvent association, US Federal LCA Commons, ENEA, NORSUS and ADEME. Among the main results, it was noted that GLAD is a very useful data aggregator or metadata catalog and nodes recommended increased emphasis on search application functionality and revision of metadata guidelines with consideration of search and cataloguing Best Management Practices.

2. Summary of descriptors

The following table contains the full list of descriptors currently available in GLAD. Data providers wishing to connect must complete the **mandatory** descriptors. **Recommended** descriptors are encouraged to improve data findability as they are linked to the fields in the faceted search, while **optional** descriptors are meant to provide users additional information so that they can assess if a particular dataset is useful for their purpose.

The columns in the following table refer to:

- Name: The name of the metadata descriptor as displayed in API / GLAD
- GLAD ID: The ID of the metadata descriptor as displayed in API / GLAD
- Displayed in UI: Indicates if the descriptor is displayed in GLAD's user interface
- Classification: defined by the WG as:
    - Mandatory: minimum requirements to connect to GLAD
    - Recommended: to enhance findability
    - Optional: desirable for fitness for purpose
- Type: Type of information required to complete the metadata descriptor. It can be one of the following:
    - String: the descriptor is a text value
    - Integer: the descriptor is a numerical value
    - Enum: the descriptor has a fixed set of values, i.e., following a predefined list of relevant alternatives
    - Boolean: the descriptor can be either true or false
- Enum values: List of predefined alternatives for the metadata descriptor (only for descriptors that are Enum type).
- Enum type: Only for descriptors that are Enum type, one of the following
    - Multiple: multiple values are allowed
    - Single: only one value is allowed
- Default value: Value automatically set for the metadata descriptor if no information is provided by the data provider (mandatory descriptors are mandatory to complete)

Table 1: Complete list of Metadata Descriptors in GLAD

Metadata Descriptor	Parameter in the API	Classification	Displayed in UI	Type	Enum values	Enum type	Comments
Unique identifier of dataset	*refId*	Mandatory		String	NA	NA	
Name of the dataset	*name*	Mandatory	x	String	NA	NA	
URL to dataset or database	*datasetUrl*	Mandatory	x	String	NA	NA	
Dataset publicly accessible?	*publiclyAccessible*	Deprecated		Boolean	NA	NA	
Category	*categories*	Mandatory	x	String	NA	NA	One entre per child category
UNSPSC process code	*unspscCode*	Deprecated		String	NA	NA	
Description of the dataset	*description*	Mandatory	x	String	NA	NA	
Technology	*technology*	Recommended	x	String	NA	NA	
Data format	*format*	Mandatory	x	Enum	ECOSPOLD1 ECOSPOLD2 ILCD JSON-LD OTHER	Single	No default value is defined. An empty entry for this field won't be permitted.

					UNKNOWN		
Regional code	*location*	Mandatory	x	String	NA	NA	
Latitude of the geography	*latitude:*	Optional		number (double)	NA	NA	
Longitude of the geography	*longitude:*	Optional		number (double)	NA	NA	
Data Provider	*dataprovider*	Mandatory	x	String	NA	NA	
Process type	*processType*	Mandatory	x	Enum	UNIT PARTIALLY_AGGREGATED FULLY_AGGREGATED BRIDGE UNKNOWN	Single	No default value is defined. An empty entry for this field won't be permitted.
LCI modeling approach	*modelingType*	Mandatory	x	Enum	ATTRIBUTIONAL CONSEQUENTIAL BEFORE_MODELING UNKNOWN	Single	No default value is defined. An empty entry for this field won't be permitted.
Multifunctional modelling	*multifunctionalModeling*	Recommended	x	Enum	PHYSICAL ECONOMIC CAUSAL SYSTEM_EXPANSION OTHER APPROACH	Multiple	Default value: UNKNOWN

					NONE UNKNOWN NOT_APPLICABLE		
Dataset contact	*contact*	Mandatory	x	String	NA	NA	
valid from year	*validFromYear*	Mandatory	x	integer (int32)	NA	NA	
valid from timestamp	*validFrom*	Optional		integer (int64)	NA	NA	
valid until year	*validUntilYear*	Recommended	x	integer (int32)	NA	NA	
valid until timestamp	*validUntil*	Optional		integer (int64)	NA	NA	
Available for free? (vs. for purchase)	*free*	Mandatory	x	Boolean	NA	NA	
Supported nomenclature	*supportedNomenclatures*	Recommended	x	String	NA	NA	
Review Type	*reviewType*	Recommended	x	Enum	INTERNAL REVIEW EXTERNAL REVIEW PANEL UNKNOWN	Multiple	Default value: UNKNOWN

					NONE		
Review system	*reviewSystem*	Optional		Enum	ILCD PEF GHG LCA_UN OTHER UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
Reviewing persons	*reviewers*	Optional	x	List of String	NA	NA	
License Type	*license*	Recommended	x	String	NA	NA	
Copyright protected	*copyrightProtected*	Optional	x	Boolean	NA	NA	
Copyright holder	*copyrightHolder*	Optional	x	String	NA	NA	
Representativeness type	*representativenessType*	Optional		Enum	SCIENTIFIC EXPERT_BASED	Multiple	Default: EXPERT_BASED
Completeness	*completeness*	Optional		Number (double)	NA	NA	Default:100
Biogenic carbon modeling	*biogenicCarbonModeling*	Optional		Enum	OMITTED DISTINGUISHED	Multiple	Default: NOT_APPLICABLE

					AGGREGATED UNKNOWN NOT_APPLICABLE		
End of life modeling	*endOfLifeModeling*	Optional		Enum	CUT_OFF PHYSICAL_APOS ECONOMIC_APOS SUBSTITUTION OTHER UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
Water modeling	*waterModeling*	Optional		Enum	AMOUNTS AMOUNTS_AND_AVAILABILITY AMOUNTS_AND_QUALITY UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
Infrastructure modeling	*infrastructureModeling*	Optional		Enum	INCLUDED_AND_DISTINGUISHED INCLUDED_AND_NOT_VISIBLE NOT_INCLUDED UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE

Emission modeling	*emissionModeling*	Optional		Enum	INCLUDED_AND_DISTINGUISHED INCLUDED_AND_NOT_VISIBLE NOT_INCLUDED UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
Carbon storage modeling	*carbonStorageModeling*	Optional		Enum	INCLUDED_AND_DISTINGUISHED_CORRECTION INCLUDED_AND_DISTINGUISHED_OTHER INCLUDED_AND_NOT_VISIBLE NOT_INCLUDED UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
Source reliability	*sourceReliability*	Optional		Enum	MEASURED_VERIFIED PARTLY_MEASURED_VERIFIED PARTLY_MEASURED_PARTLY_ESTIMATED ESTIMATED_QUALIFIED ESTIMATED_UNQUALIFIED	Multiple	Default: NOT_APPLICABLE

Aggregation type	*aggregationType*	Optional		Enum	HORIZONTAL VERTICAL COMBINED UNKNOWN NOT_APPLICABLE	Multiple	Default: NOT_APPLICABLE
CO2PE product code	*co2peCode*	Optional		String	NA	NA	
supported LCIA methods	*lciaMethods*	Optional		List of String	NA	NA	
Representativeness	*representativenessValue*	Optional		Double	NA	NA	
Deviation in mass and energy balance	*amountDeviation*	Optional		Double	NA	NA	

3. Guidance on Mandatory Descriptors

Unique identifier of dataset

General guidance related to UI:

This field provides the code used by the data provider to identify a particular dataset. It is not shown in GLAD's user interface.

Guidance related to API for data providers:

Data providers should complete the parameter refId using a universally unique identifier (UUID)[4] for each dataset. The UUID is normally generated by the LCA of database software in which the dataset is developed.

Name of the dataset

General guidance related to UI:

This field acts as the title of the data set. It is provided as the main result when performing a search in GLAD.

Guidance related to API for data providers:

Data providers should complete the parameter name with a free text that provides general identification of the dataset. No specific naming convention is required, however, it is recommended that this field includes the following components (based on the ILCD Handbook[5]):

- A general descriptive name of the process, using technical language, as it is used in the respective industry or towards their customers.

[4] https://en.wikipedia.org/wiki/Universally_unique_identifier

[5] European Commission - Joint Research Centre - Institute for Environment and Sustainability: International Reference Life Cycle Data System (ILCD) Handbook - Nomenclature and other conventions. First edition 2010. EUR 24384 EN. Luxembourg. Publications Office of the European Union; 2010"

- Qualitative information such as treatment received (e.g. "polished", "cleaned", "chromium plated", "sterilised", etc.), standard fulfilled (such as for material grades/purity, fulfilled emission limits, etc.), product quality (e.g. "glossy", "UV-resistant", "flame-retardant", "antibacterial finishing", etc.), use information (e.g. "indoor use", "bottle grade", "for wafer production", etc.), production route ("suspension polymerisation", "spray dried", "Fischer-Tropsch", etc.), etc.
- Whether "Production mix" (weighted average mix of production routes) or "Consumption mix" (weighted contribution of imported and exported products)
- Information about the location type of availability such as: "at plant", "at wholesale", "at point-of-sale", "to consumer", etc.
- Quantitative specifying information such as qualifying constituent(s)-content and / or energy-content per unit, etc.

URL to dataset or database

General guidance related to UI:

This field provides the URL address where users have access to the data. Users will find the “Go to Dataset” option when data providers provide a single URL to a particular dataset or database. On the other hand, if datasets are not publicly available to download from the internet, users will find a “Go to website” option in GLAD, that will direct them to the data provider’s website to make the relevant enquiries about the data required.

Please beware that this is under implementation and users may still find the “Go to Dataset” option that is still linked to the data provider’s corporate website rather than the dataset or database.

Guidance related to API for data providers:

Data providers should complete the parameter dataSetUrl with the URL to the specific single dataset. If the data provider does not offer the possibility to access and download data online, this field should be left blank.

In all cases, the URL of the corporate website or landing page of the data provider (like specified in the user account) will be included in the dataset information as “Go to website”.

There is another metadata descriptor “Dataset publicly accessible?” to indicate whether the dataset is open and available to download somewhere on the internet, regardless of whether it is free or for purchase. The WG decided this info was replicated in the descriptor “URL to

dataset or database". Thus, the parameter publiclyAccessible might be removed in the future. Data providers can choose not to complete this parameter.

Category

General guidance related to UI:

This field aims to classify datasets into categories so that they can be easily or intuitively found. It is used in GLAD as a search filter and users may narrow down their results selecting one of the available options. Beware that GLAD currently does not require any particular standardised classification system (see definition below) and therefore, the current category facet does not provide a unique and consistent classification tree with parent-child relationships. Instead, the facet provides a very extensive list of all categories that are used by nodes, which may or may not follow a standardised system. For this reason, the facet allows users to type in the keyword for the category and to select multiple categories in a single search.

It must be noted that a new feature is under development that will allow for internal mapping of any category set by data providers onto ISIC, to facilitate search. ISIC is GLAD's preferred classification system.

Guidance related to API for data providers:

Data providers should complete the parameter categories introducing an individual entry per child category of the dataset as an array. e.g. categories = ['Agriculture', 'Corn Production'] . With this information, the parameter category will be automatically built from the elements in the 'categories' field, concatenated with a slash (/). e.g. categories = ['Agriculture', 'Corn Production'] → category = 'Agriculture/Corn Production'

GLAD currently does not require any particular standardised classification system (see definition below), so any category term will be accepted. However, the preferred category system for GLAD is ISIC and its use is highly encouraged and recommended to improve consistency. Data providers should complete parameter categories as an array introducing an individual entry per child category including code number, e.g. categories = ['A. Agriculture, forestry and fishing', '01. Crop and animal production, hunting and related service activities', '011. Growing of non-perennial crops']. If ISIC is not used, any other standardised process/activity classification system is recommended (see definition below).

It must also be noted that currently, the name of the classification system cannot be provided, but a new feature is under development to allow that to help make the supplied categories more consistent.

It must also be noted that a new feature is under development that will allow for internal mapping of any category set by data providers onto ISIC. This feature will be directly available for datasets using any standardised process classification. Please beware that mapping between process/activity and product classification may possess several restrictions due to multifunctionality. Nodes are encouraged to provide any additional information about classification in the "Description" field.

An additional parameter is available in the API unspscCode to provide coding in United Nations Standard Products and Services Codes (UNSPC) System. This might be eliminated in the future. Data providers can choose not to complete this parameter.

Definition

Categorization is usually used in modelling software as a means of organising and managing datasets so that they can be easily or intuitively found and used for building a particular product system or model. Standardised classification systems are hierarchical structures. They are usually structured as parent-child relationships in which processes and flows are organised to demonstrate relationships to each other. Category labels and names are therefore an important part of dataset documentation or metadata that exist in datasets and provide an opportunity to organise data so that they might be extracted by a search application for effective search and retrieval.

Categorization can take place at the process or activity level, that is based on the unit processes included in the dataset, or at the product level, that is, based on the output products of the unit processes covered by the dataset.

Standard classification system for process/activity categorization include:

- **International Standard Industrial Classification (ISIC)[6]:** it is the international reference classification of productive activities, developed by UN Statistics division, that is widely used by the majority of countries around the world as their national activity classification or have developed national classifications derived from ISIC, relevant for comparing statistical data on economic activities at the international level. The most recent version is 4.0

[6] https://unstats.un.org/unsd/classifications/Econ/isic

Example of levels:

- Section A: Agriculture, forestry and fishing
- Division 01 Crop and animal production, hunting and related service activities
- Group 011 Growing of non-perennial crops
- Class 0111 Growing of cereals (except rice), leguminous crops and oil seeds

- **North American Industrial Classification System (NAICS)[7]:** it is the standard used by US Federal statistical agencies to classify business establishments for the purpose of collecting, analysing, and publishing statistical data. It has been developed in coordination with authorities in Canada and Mexico to allow for a high level of comparability in business statistics among the North American countries. It replaces the previous Standard Industrial Classification (SIC) system. The most recent version is for year 2022

  Example of levels:

  - Sector 11 Agriculture, Forestry, Fishing and Hunting
  - Subsector 111 Crop Production Crop Production
  - Group 1111 Oilseed and Grain Farming
  - Industry 11111 Soybean Farming

Standard classification system for product categorization include:

- Central Product Classification (CPC)[8]: it is the international reference classification structure for products (goods and services) used for assembling and tabulating all kinds of data requiring product detail, including statistics on industrial production, domestic and foreign commodity trade, international trade in services, balance of payments, consumption and price statistics and other data used within the national accounts. It is developed by the UN Statistics division and the current version is 2.1.
- United Nations Standard Products and Services Codes (UNSPSC)[9]: developed by the UN Development Programme (UNDP), as an open, global, multi-sector standard for efficient, accurate classification of products and services.

[7] https://www.census.gov/naics/reference_files_tools/2022_NAICS_Manual.pdf
[8] https://unstats.un.org/unsd/classifications/Econ/cpc
[9] https://www.unspsc.org/

- Harmonized System (HS)[10]: developed by the World Customs Organization for the import and export classification systems used in the United States and by many trading partners.
- North American Product Classification System (NAPCS)[11]

The LCA community does not use a standard or even consistent classification schemes for process and flow categorisation. In some cases, datasets are categorised by the type of process or activity it covers, such as ecoinvent that uses ISIC system, and the United States Federal LCA Commons uses NAICS. In others, data providers categorise their databases according to a customised system based on standard, such as IDEA that uses a tailored system based on Japan Standard Industrial Classification System.

Individual and institutional researchers usually use their own categorisation approach or perhaps a mix of categorization schemes if they import heterogeneous data sets into their software. For example, if a researcher imports ecoinvent data into their software, the ecoinvent data will be categorised according to the ISIC system. The researcher might continue to use this scheme and/or build and organise process models outside this scheme. Therefore, when these data sets are exported, the underlying files will have inconsistent categorisation schemes. Further, organisations that provide commercial or open data products, may also use classification schemes that are not consistent with each other. This represents a missed opportunity to provide a consistent convention that might be used as a powerful search tool, but also to build capacity for more advanced modelling and data integration approaches that are starting to be developed

Description

General guidance related to UI:

This field provides a description of the general scope of the dataset. Its first few lines are provided together with the “Process Name” of the dataset as results when performing a search in GLAD. The full content of this descriptor can be seen in GLAD’s user interface when the “more” option is selected.

[10] https://www.trade.gov/harmonized-system-hs-codes
[11] https://www.census.gov/naics/napcs/?8976654

Guidance related to API for data providers:

Data providers should complete parameter description with a free text that fully describes the dataset. It is recommended that it contains a combination of the following:

1. An introduction of the technology description and PROCESS DESIGN paragraph
2. The boundary conditions to indicate general aspects of the system boundaries
3. Any additional details key to understand the process

The information of this field may or may not be a repeat of metadata included in other fields.

The following is an example of the Description field for the dataset "Calcium carbonate, ground, 20 micron, at plant":

> This process represents the production of "Calcium carbonate, ground, 20 micron, at plant" using average technologies for the United States from 2015-2016. The process includes three sub-processes: Quarry Operations; Transport and Plant Processing. Quarry Operations includes the following unit operations: mechanical extraction; primary crushing; screening; and intermediate storage of calcium carbonate rock (marble, limestone, or chalk). Transport includes the transport of materials from Quarry Operations to Plant
>
> Processing via barge, train, or truck. Plant processing which includes jaw crushing, washing, impact crushing, ball milling to particle size, and then classifying.
>
> The system boundary includes: 1) the transport of raw materials to multiple manufacturing facilities where various subcomponents are produced; 2) the manufacture of subcomponents; 3) the transport of subcomponents to a different manufacturing plant for final assembly; 4) the assembly of subcomponents into a complete scanner; and 5) the transport of generated waste from the manufacturing facilities to a municipal solid waste landfill. The following processes and life cycle phases fall outside the system boundary: 1) packaging of the completed scanner; 2) all transport downstream of the assembly plant gate; 3) sale of product; 4) product use phase; and 5) end-of-life phase (including recycling).

Data format

General guidance related to UI:

This field provides information about the standardised format in which the dataset is available. It is used in GLAD as a search filter, and users may narrow down results using any of the following:

- ECOSPOLD1: the data format is version 1 of "ecoSpold" (see definition below)
- ECOSPOLD2: the data format is version 2 of "ecoSpold" (see definition below)
- ILCD: the data format is "ILCD" (see definition below)
- JSON LD: the data format is "JSON LD" (see definition below)
- OTHER: the data format is any other different from the above

Guidance related to API for data providers:

Standardised data formats enable data exchange. ISO/TS 14048:2002 provides the requirements and a structure for a data documentation format, to be used for transparent and unambiguous documentation and exchange of LCA and LCI. Data providers should complete the parameter format selecting a single one of the following options:

a. ECOSPOLD1: for datasets that were developed using the version 1 of "ecoSpold" format (see definition below)
b. ECOSPOLD2: for datasets that were developed using the version 2 of "ecoSpold" format (see definition below)
c. ILCD: for datasets that where developed using the "ILCD" format (see definition below)
d. JSON-LD: for datasets that where developed using the "JSON-LD" format (see definition below)
e. OTHER: for datasets that were developed using any other format
f. UNKNOWN: this alternative is inconsistent and will eventually be removed. Therefore, data providers must not use this option. If data sets were developed with a format different to alternatives a, b, c, or d, then "OTHER" should always be selected.

Please beware that an empty entry for this field won't be permitted.

Definitions:

ecoSpold: is primarily associated primarily with the ecoinvent LCI database, however, it is an open-source format, based on XML. It is available in version ecoSpold1 (used in

ecoinvent up until ecoinvent v2.2) and its most recent version, ecoSpold2 (Meinshausen et al. 2016). The ecoSpold format is supported by most major LCA software applications, and it is used, besides for the ecoinvent database, by the AusLCI in Australia, PeruLCA in Peru and the Quebec LCI database in Canada.

ILCD: is the data format used for the European Reference Life Cycle Database (ELCD) and for the development of data under the Environmental Footprint scheme (PEF/OEF). It is also XML-based format. It is supported by most major LCA software applications and it is used by national LCA databases, such as SICV in Brazil, MYLCID in Malaysia and the Thai National LCI database.

JSON-LD: A technical alternative to XML that combines ILCD and ecoSpold. JSON-LD is a format based on JavaScript Object Notation for Linked Data (JSON-LD) developed by GreenDelta. Besides the aim to reduce the effort for implementation and remove inconsistencies between the ILCD and ecoSpold formats, there are other advantages, such as being human-readable and the ease of integration into web-applications. JSON-LD was implemented as one of the formats used by openLCA in 2015, but use of this format in other LCA software is still limited.

More information can be found at
https://helpdesk.lifecycleinitiative.org/distribution/data-formats/

Regional code

General guidance related to UI:

This field contains information about the geographical validity of the dataset. It is used in GLAD as the search filter “Geographical coverage”, and users may narrow down their results selecting one of the available options. Please beware that not a particular convention is used for geographical classification.

Guidance related to API for data providers:

Data providers should complete the parameter location providing information about the geographical scope with the level of aggregation relevant to the dataset (world region, country, region within a country, etc). Not a particular convention is used for this field, however, it is recommended that data providers use the 2-letter geographical classification of ISO 3166, with one or two levels of classification, such as AR for Argentina and BR-MG for the state of Minas

Gerais in Brazil. GLAD keeps a file online with a list of recommended codes for location based on this standard, with some additions of 3-letter regional codes such as RSA (Region South America), RER (Europe), GLO (Global), RoW (Rest of World) etc. The use of this list is recommended.

It is acknowledged that the developed list mainly fits countries or regions in the administrative level and it may not fit all required situations where a particular dataset has a geographical coverage beyond administrative limits. Therefore, a new metadata descriptor such as "Geographical representativeness" is being considered in order to provide additional information as a free text to describe how well the dataset fits the regional code defined in this descriptor.

Please beware that there are two parameters available in the API latitude and longitude which are not currently being used but might be relevant in future development to link shape files as metadata for location of the data. For the time being, nodes may complete these two parameters optionally.

Data provider

General guidance related to UI:

The name of the institution that submits datasets to GLAD (In GLAD, the data provider is also called Node). It is used in GLAD as the search filter "Data Provider", and users may narrow down their results selecting one of the available options.

Guidance related to API for data providers:

The parameter dataprovider is completed automatically when the node submits its data. It should be noted that data provider does not refer to the data generator, who is the person or organisation responsible for the modelling of the process and the compilation or the updating of the data (ISO/TS 14048:2002). Please refer to the mandatory descriptor "Dataset contact" for more details.

It is acknowledged that additional metadata descriptors are required in order to provide information about:

a) *versioning: data providers with several versions of a database due to updates of data and other adjustments.*

b) *data stocks: data providers that possess bundles of datasets with particular characteristics that need to be used together to describe the system.*

A new descriptor "Collection" is being discussed for this purpose. However, it is not yet implemented. Therefore, data providers are required to provide any required information in the description field.

Process type

General guidance related to UI:

This field indicates the type of process covered by the dataset. It is used in GLAD as a search filter, and users may narrow down results using any of the following:

- Unit: the process type is "Unit process" (see definition below)
- Fully aggregated: the process type is "Fully aggregated process" (see definition below)
- Partially aggregated: : the process type is "Partly aggregated process" (see definition below)

Guidance related to API for data providers:

Data providers must specify the parameter processType selecting a single one of the following options:

- UNIT: for datasets whose process type is "Unit process" (see definition below)
- PARTIALLY_AGGREGATED: for datasets whose process type is "Partly aggregated process" (see definition below)
- FULLY_AGGREGATED: for datasets whose process type is "Fully aggregated process" (see definition below)
- BRIDGE: this type of process type is deprecated and will be eventually removed. Therefore, we request that data providers don't use this option. The definition is provided below.
- UNKNOWN: this value can be used when no information is known about the type of process. However, it is highly recommended that data providers avoid using this option for this metadata descriptor.

Please beware that an empty entry for this field won't be permitted.

Definitions:

Unit process: According to ISO 14044:2006, a unit process is the "smallest element considered in the life cycle inventory analysis for which input and output data are quantified". This term is applied both to processes which cannot be further subdivided and to a process-chain, plant- or facility-level scope (e.g., a petroleum refinery). The ILCD Handbook (2010) designates these two different kinds of unit processes as "unit process, single operation" and "unit process, black box", respectively. A "unit process, black box" encompasses more than one "unit process, single operation". The central feature of a unit process is that only direct exchanges (inputs/outputs) with the natural environment are considered as elementary flows in the unit process inventory. All other inputs and outputs are consequently product (i.e., reference or intermediate) flows exchanged between human activities.

Partly aggregated process: this implies that at least one or more input/output flows requires further modelling, i.e., to be connected (linked) to other processes, to complete the life cycle inventory (LCI) of the product system. This dataset scope offers flexibility to change/adapt specific aspects, e.g., electricity input or waste treatment scenario, while the rest of the product system is pre-determined/defined. This corresponds to the terms "partly terminated aggregated process dataset" (UNEP 2011) and "partly terminated system" (EC-JRC 2020).

Fully aggregated process: Also known as "fully terminated aggregated process dataset (UNEP 2011) or "LCI results" (EC-JRC 2020). This dataset only comprises an entire product system (as either a complete life cycle or parts thereof) and only has elementary flows, besides the reference product(s), crossing its system boundary.

Bridge processes: A bridge process is a process that connects a unit process from one dataset with a CUTOFF, to a process, in another dataset or database to provide a given technosphere product. Bridge processes (BPs) are constructs that do not describe any original activities but only serve to create connections to enable full life cycle modelling[12]. Please beware that this term is deprecated and will be removed in future versions of the API. Therefore, we request that data providers don't use this option.

[12] Ingwersen, W.W., Kahn, E. & Cooper, J. Bridge processes: a solution for LCI datasets independent of background databases. *Int J Life Cycle Assess* 23, 2266–2270 (2018). https://doi.org/10.1007/s11367-018-1448-6

LCI modelling approach

General guidance related to UI:

This field provides information about the type of life cycle inventory modelling approach used for the dataset. It is used in GLAD as a search filter, and users may narrow down results using any of the following:

- Attributional: the dataset was modelled using an "attributional approach" (see definition below)
- Consequential: the dataset was modelled using a "consequential approach" (see definition below)
- Before modelling: the dataset is presented before modelling (see definition below)

Guidance related to API for data providers:

Data providers should complete the parameter modelingType selecting a single one of the following options:

- ATTRIBUTIONAL: for dataset where allocation was performed using an "attributional approach" (see definition below)
- CONSEQUENTIAL: for dataset where allocation was performed using a "consequential approach" (see definition below)
- BEFORE_MODELING: for dataset where no allocation was performed (see definition below)
- UNKNOWN: this value can be used when no information is known about the modelling approach. However, it is highly recommended that data providers avoid using this option for this metadata descriptor.

Please beware that an empty entry for this field won't be permitted.

Definitions:

> **Attributional approach[13]:** 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.

[13] United Nations Environment Programme, 2011. Global Guidance Principles for Life Cycle Assessment Databases.

Consequential approach: 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.

Before modelling: No modelling approach has been applied to deal with multi-functionality - nor linking to (average/marginal) supplying or demanding activities has been established. For example, in version 3 of the ecoinvent database, this corresponds to the "undefined" version of datasets, i.e., prior to any system modelling.

Dataset contact

General guidance related to UI:

This field provides the contact information of the data generator, that is, the institution or person that collected the information and developed the dataset. It can be seen in GLAD's user interface when the "more" option is selected.

Guidance related to API for data providers:

Data providers should complete the field contact with the name of the author and/or organisation that generated the data. Generic contact, such as website or helpdesk, can be also added here.

This field acknowledges the data generator in order to be consistent with the requirements in the FAIR principles for data. However, it is the data provider's (node) responsibility to secure consent of the data generator for the publication of their contact details. Special attention must be given to the disclosure of personal information such as names, email address, telephone, etc. of individuals.

If no information of the data generator is available or meant to be disclosed, the contact details of the data provider should be given in this field.

Valid from

General guidance related to UI:

This field contains a year value to refer to the time validity of the dataset. It is used in GLAD as a search filter, and users may select a year value for "valid from" and "valid until" to narrow down the Reference year of the dataset.

Guidance related to API for data providers:

Data providers should complete at least one of the two parameters regarding this descriptor:

- validFrom: value in Unix timestamp[14] format (milliseconds since 01/01/1970)
- validFromYear: year value in YYYY format (it will be taken from validFrom if not set)

This value can either be the starting point of validity period (valid from), in which case the field "valid until" shall be also completed. If this is not the case, the "valid from" value will be considered as the reference year of the dataset (as "valid for").

Available for free? (vs. for purchase)

General guidance related to UI:

This field indicates whether the dataset is available for free or for purchase, according to its licensing terms and conditions. It is used in GLAD as a search filter, and users may narrow down selecting either "free" or "for sale". Please beware that in either case, users will be redirected to the data provider's website to download the dataset and this may require registration even though the datasets are available for free.

Guidance related to API for data providers:

This descriptor only permits the values "true" or "false". Data providers should complete the parameter free using "true" when data is available free of charge and "false" otherwise (e.g. when a paywall appears when trying to download data).

[14] https://www.unixtimestamp.com/

Please note that a dataset can be free under the condition that one has a valid background data licence. Licensing terms can be described under the descriptor “Dataset licence”.

4. Guidance for Recommended Descriptors

Technology

General guidance related to UI:

This field gives a short description on the technology used in the process, which is modelled in the dataset. This helps data users to evaluate whether this dataset (technology) can be used in a specific LCA study. It can be seen in GLAD's user interface when the "more" option is selected.

Guidance related to API for data providers:

Data providers should complete the parameter technology with a string that describes the technology used in the dataset. Overlap with the "description" field is possible. However, the technology field allows for more detailed information on the technology used.

Supported Nomenclatures

General guidance related to UI:

Information about the compliance of the dataset with one or multiple of the nomenclature systems for elementary flows (see definition below). It is used in GLAD as a search filter and users may narrow down their results selecting one of the available options. Please beware that not a particular convention is used for nomenclature classification.

Guidance related to API for data providers:

Data providers should complete the parameter supportedNomenclatures with information about the main Nomenclature System in which the dataset was created (see definitions below). GLAD offers the possibility of data conversion among different formats. Since conversion might result in data loss, please provide the nomenclature system in which the dataset is created.

Beware that GLAD does not require any specific convention for nomenclature classification. Instead, the following recommendations should be followed by data providers when completing this field:

- Include the name and version of the standard nomenclature system. e.g: ecoinvent 3.6. Main standard nomenclature systems are described in the definitions section below.
- Avoid using the name of the format to refer to the name of the nomenclature system. e.g. ECOSPOLD2 is not a nomenclature system
- If the dataset is created in commercial LCA software, indicate the name of the software and version
- If new elementary flows were created in a standardised system, write “Customised” for this field and include more information in the description field
- Avoid using the word “other” without further information.
- Keep the wording short, as this field is used as a facet for searching purposes.

It is highly recommended that elementary flows which were created or that deviate from the original nomenclature system are detailed under the mandatory descriptor “Description”. In this way, the user has the possibility to evaluate whether all elementary flows included by the dataset generator can be characterised when reusing the dataset.

Definitions:

Nomenclature systems include the naming structure of elementary flows and are essential during impact assessment because characterization in LCIA methods depends on this convention. There are four main nomenclature systems that are linked to the format of datasets. These are:

- ecoinvent
- EF/ILCD
- IDEA
- FEDEVL

Each of these are regularly updated and can be found in several version numbers. Please refer to the develop website for more information. Besides these, LCA software tends to operate with a customised nomenclature system that merges nomenclature systems from different background databases and aligns these with available LCIA methods. In addition, LCA

practitioners can also define new elementary flows that align with the scope of the study (for example related to plastic littering).

To date, some of the commonly systems that were indicated by nodes in this field are:

- EF3.1
- EF3.0
- EF2.0
- ILCD
- ecoinvent 3.6
- ecoinvent 3.5
- ecoinvent 2.2
- IDEA v2.3
- FEDEVL v1.0.3.
- GaBi

Multifunctional modelling

General guidance related to UI:

This field provides information about the method used to deal with multifunctional processes. It can be seen in GLAD's user interface when the "more" option is selected.

Guidance related to API for data providers:

Data providers should complete the parameter multifunctionalModeling selecting a one of the following options:

- PHYSICAL: for dataset that used physical allocation to deal with multifunctional process (see definition below)
- ECONOMIC: for dataset that used economic allocation to deal with multifunctional process (see definition below)
- CAUSAL: this type of modelling is deprecated and will be eventually removed. Therefore, we request that data providers don't use this option.
- SYSTEM_EXPANSION: for dataset that used system expansion to deal with multifunctional process (see definition below)

- OTHER APPROACH: for datasets that use any other approach to deal with multifunctional processes. Further information about this approach should be provided in the “Description” field.
- NONE: for datasets that are strictly single-output (mono-functional)
- NOT_APPLICABLE: this value should be used only when the field “LCI modelling approach” is set as "Before modelling"
- UNKNOWN: this value can be used when no information is known about the modelling approach. However, it is highly recommended that data providers avoid using this option for this metadata descriptor. Please beware that if no information is provided, this is the default value for this field

This field is dependent on the metadata descriptor "LCI modelling approach". For the sake of interoperability, it is highly recommended that data providers complete this field when the LCI modelling approach is other than "Before modelling". Please consider that not providing this information possesses the risk of unintentional misuse and lack of interoperability.

Definitions:

Physical allocation[15]: partitioning of multifunctional processes is based on physical properties such as mass, energy content, exergy content, or concentration

Economic allocation[16]**:** partitioning is based on the economic value such as market prices of products and services, of primary materials, recycled materials, or scrap

System expansion: this approach solves the allocation problem by expanding the original system boundaries of the system being studied. In a comparative setting, this might entail including additional functions to ensure that two or more options provide the same function(s). This (additive) approach is of limited relevance for LCI databases. Instead, the system boundaries might be expanded to also consider substitutes for ‘non-reference’ co-products, with “credits” (i.e., avoided burdens) assigned to the reference product based on alternative routes for providing these functions. In this context, ”substitution is a special (subtractive) case of applying the system expansion principle.” (UNEP 2011 Shonan Guidelines)

[15] United Nations Environment Programme, 2011. Global Guidance Principles for Life Cycle Assessment Databases"

[16] United Nations Environment Programme, 2011. Global Guidance Principles for Life Cycle Assessment Databases"

Review Type

General guidance related to UI:

This field provides information about the type of review to which the dataset was submitted. It is used in GLAD as a search filter, and users may narrow down results using any of the following:

- External: the dataset have gone through an external review (see definition below)
- Internal: the dataset have gone through an internal review (see definition below)
- Panel: the dataset have gone through a review process with a panel (see definition below)

Guidance related to API for data providers:

Dataset providers should complete the parameter reviewType selecting a single options from the following:

- INTERNAL: for dataset that have gone through an internal review process (see definition below)
- EXTERNAL: for dataset that have gone through an external review process (see definition below)
- PANEL: for dataset that have gone through review by a panel (see definition below)
- NONE: for datasets that haven't gone through any type of review
- UNKNOWN: this value can be used when no information is known about the review of the dataset. However, it is highly recommended that data providers avoid using this option for this metadata descriptor. Please beware that if no information is provided, this is the default value for this field

Please note that other metadata descriptors are optional to provide further information about the review process. reviewSystem should be completed to indicate whether a particular review system was used. reviewers should be completed with the name of the person(s) that performed the review. Nodes are encouraged to provide any additional information about classification in the "Description" field.

It is acknowledged that in order to be compliant with different reporting schemes, datasets should fulfill a set of requirements which include data format, nomenclature, review system, etc. A new metadata descriptor "Compliance" is being considered in order to provide information about whether data can be used in a particular scheme. This is not currently available in GLAD

Definition:

Critical review is the process intended to ensure consistency with the principles and requirements defined in the standard (ISO14040). It could be performed externally by a third-party, independent to the data owner, or internally by the data owner. A Panel Review is that which is compliant to ISO 14040, where a panel of 3 external reviewers is required.

License type

General guidance related to UI:

This field provides a description about the terms and restrictions for data access and use (see definition below). It is provided in GLAD under the “more details” options of a dataset.

Guidance related to API for data providers:

Data providers should complete the parameter license with information about the type of license that sets the terms and restrictions for data access and use (see definition below). The following minimum information and recommendations should be followed by data providers when completing this field:

- Indicate whether it is an open or restricted license
- If it is an open license, indicate if it is one of the common licenses described in the definition section below. If not, provide information about the terms of the open license.
- If it is a restricted license, indicate how to obtain the EULA form the copyright holder
- Include any other relevant information about required licenses, for example, for background data used in the dataset.

Please beware that the type of license is different to whether the data is available for free or not. That information can be found in the mandatory descriptor “available for free?”

Also, information about the copyright of the data is not provided in this field. An optional descriptor copyrightProtected exists to indicate if the dataset is copyright protected (see definition below) and copyrightHolder to indicate the name of the owner of the copyright of the database. The creator of the dataset is acknowledge under “dataset contact”

Definitions:

Copyright refers to the legal protection that authors of creative works (in this case, datasets) get in order to that prevent the use of their work by others without permission[17]. It is related to the owner of the dataset. In order to enable access and use by others of the datasets, owners or copyright holders can set a license that describes the terms and conditions for the reuse of data. In contrast to copyright protected, works can also be in the public domain, that is not subject to copyright protection, such as very old works or works that have been released to the public domain by their authors

Licences can bradley be classified into:

Open License[18]: a license that allows users to use a resource without seeking permission from the owner. This type of license is used by authors who want to share their work openly but still want to maintain some control over what others can do with it. Some common template for open licenses include:

- Creative Commons[19]
- GNU General Public License (GNU-GPL)[20]
- MIT License[21]

Restricted license: the access and use of the works is defined in a particular End-User License Agreement (EULA) set by the owner. This is the most common licence for commercial databases.

Valid until

General guidance related to UI:

This field contains a year value to refer to the time validity of the dataset. It is used in GLAD as a search filter, and users may select a year value for “valid from” and “valid until” to narrow down the Reference year of the dataset.

[17] https://www.copyright.gov/what-is-copyright/
[18] https://edtechbooks.org/k12handbook/copyright
[19] https://creativecommons.org/
[20] https://www.gnu.org/licenses/gpl-3.0.html
[21] https://opensource.org/licenses/MIT

Guidance related to API for data providers:

It refers to the end point of the validity period (valid until). Data providers should complete at least one of the two parameters regarding this descriptor:

- validUntil: value in Unix timestamp[22] format (milliseconds since 01/01/1970)
- validUntilYear: year value in YYYY format (it will be taken from validUntil if not set)

5. Remarks about Optional Descriptors

This guidance is not meant to provide further guidance on the optional descriptors that have been described in previous guidance. However, the following table, summarises the definition for each of these descriptors, exactly as it is provided in the document "Task 3: Core meta-data descriptors and guidance on populating descriptors" developed by WG3, published in 2017.

Name	Description
Representativeness type	*For an LCA dataset, information will typically vary by technology, producer, region, and time. This criterion assesses how representative the information is regarding these four aspects. The most representative data is that that is obtained by random sampling, although this sampling is not common practice in LCA. This field provides measurement for sample conformance.* *For science-based sampling, the coefficient of variation of the sample should be used. For expert-based sampling, the proposed measure is expert judgement. It is not recommended to use parameters such as market share or similar to guess the representativeness of the sample since these are misleading" (Ciroth et al. 2016, sample conformance)*
Completeness	*This field assesses whether the dataset is able to provide information on the main drivers for LCIA results in line with goal and scope, that is: elementary flows and processes, in case the dataset is an aggregated dataset.*
Biogenic carbon modeling	*This field described the approach for dealing with biogenic carbon modelling. 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 possible, because not all biogenic carbon emissions are reported or when biogenic flows are given different importance in the LCIA.*
End of life modeling	*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*

[22] https://www.unixtimestamp.com/

	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*
Water modeling	*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*
Infrastructure modeling	*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*

Emission modeling	*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 (ECJRC, 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*
Carbon storage modeling	*Background:* *Carbon sequestration and temporary storage is often discussed as a means to mitigate climate change, whether a bio-based or fossilbased 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*
Source reliability	*Definition:* *Reliability of the information provided in the dataset, assessed by reliability of the sources used for obtaining the information*
Aggregation type	*Definition:* *For an aggregated dataset, specify how the aggregation was performed.* *Example:*

	Horizontal, along the supply chain / vertical, across several processes delivering similar products, mixed; partial or complete – so, 6 cases)
CO2PE product code	*co2peCode*
supported LCIA methods	*lciaMethods* *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*
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*
Deviation in mass and energy balance	*amountDeviation*