

# Chapter 6

## Life Cycle Inventory Data and Databases



Andreas Ciroth and Salwa Burhan

**Abstract** Life cycle inventory (LCI) databases are commonly used in life cycle assessment (LCA) studies. They enable modern, larger case studies, make data collection more efficient, and help to establish comparability across different case studies. A database typically tries to provide one coherent and consistent modeling space, thereby allowing users to take different datasets in the appropriate database, which implies that the goal and scope of datasets in the database match the goal and scope of case studies done with the database.

This chapter explains the principal elements of LCI data, different types of databases in LCA, and explores common issues in modern LCA databases: starting a database, maintaining it, providing quality assurance, and not the least, making the database available to users. The second part of the chapter deals with data exchange and data exchange formats, as well as with interoperability concepts to allow the use of datasets from different databases in one study.

**Keywords** Background system · Data exchange formats · Data exchange · ELCD (European Reference life cycle database) · Elements of LCA data · Environmental footprint · Fitness for purpose · Foreground system · GLAD (Global Life Cycle Access to Data) · Harmonized data sets · Interoperability · LCI data · LCI databases · Life cycle assessment (LCA) · Life cycle impact assessment (LCIA) methods · Life cycle inventories (LCI) · ProBas (LCI library from the German Environmental Protection Agency) · Unified modeling language (UML) · US EPA (US Environmental Protection Agency)

---

A. Ciroth (✉) · S. Burhan  
GreenDelta GmbH, Berlin, Germany  
e-mail: [ciroth@grendelta.com](mailto:ciroth@grendelta.com); [burhan@grendelta.com](mailto:burhan@grendelta.com)

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021  
A. Ciroth, R. Arvidsson (eds.), *Life Cycle Inventory Analysis*, LCA Compendium – The Complete World of Life Cycle Assessment, [https://doi.org/10.1007/978-3-030-62270-1\\_6](https://doi.org/10.1007/978-3-030-62270-1_6)

## 1 Life Cycle Inventory Data and Databases, Definition and Introduction

As introduced in Chaps. 1<sup>1</sup> and 3<sup>2</sup> in this book, a life cycle inventory is a model of the life cycle of a product or service, with quantified inputs and outputs, and thereby comprises processes, flows, and units. When translating this into an IT model, flow properties such as mass or energy may be added, and the entire life cycle model may be called a product system, as a model of the connected processes.

Figure 6.1 shows an example from an early version of the openLCA software, where the main elements of inventory data are shown with their relations, using unified modeling language (UML) notation.

A process (dataset) is linked to one or many actors (authors, reviewers, distributors, and so forth), can contain references to one or many flows, and link to one or many sources; a flow links to one or many flow properties. A flow property refers to exactly one unit group (mass; unit groups of mass, containing, e.g., kilogram as one unit). Strictly speaking not part of the inventory are obviously life cycle impact assessment (LCIA) methods, where each method links to at least one flow. On a higher modeling level, there are product systems as structured collections of processes; one product system can contain one or several processes. Projects as comparisons of product systems are not common to all LCA data structures but they exist; evidently, then, one project contains one or more product systems.

All the elements in Fig. 6.1, apart from the LCIA methods, belong to the inventory. Therefore, all data found in and provided by these elements are inventory data.

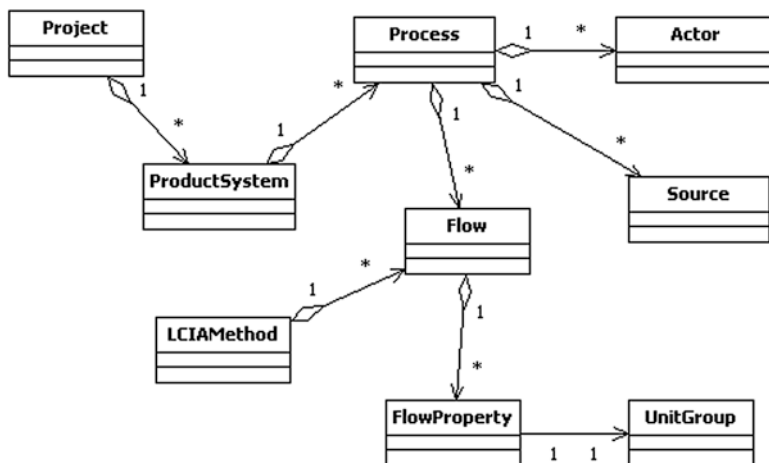


Fig. 6.1 Structure of main elements in LCA (Srocka 2009)

<sup>1</sup>Introduction to “Life Cycle Inventory Analysis”

<sup>2</sup>Development of Unit Process Datasets

Some of the elements, like the flow properties, are rather simple and do not contain a lot of more detailed information. Others, such as the process data set, are more complex and may contain, depending on the specific database or exchange format, hundreds of sub-elements, including flows with exchanges and direction, amount and unit, as well as metadata with modeling details.

A life cycle inventory (LCI) database can be defined as (UNEP/SETAC 2011, p 137):

A system intended to organize, store, and retrieve large amounts of digital LCI datasets easily. It consists of an organized collection of LCI 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 assessments and life cycle impact assessments.

The main distinction from a dataset library is the intent to provide harmonized process datasets which can be easily and without major mistakes used together, for the creation of LCA and LCIA models and for calculating them.<sup>3</sup> The definition recognizes that the attempt to provide such a harmonized, “safe” space is often not fully possible.

One of the largest, global attempts for the harmonization of process datasets and databases “culminated” in a workshop on Global Guidance Principles for Life Cycle Assessment Databases, held in Shonan, Japan in 2011, after longer preparation. These guidance principles, commonly called the “Shonan Guidance Principles,” focused on principles for creating, managing, and disseminating datasets to aid life cycle assessments of products and services globally (UNEP/SETAC 2011).

## 2 The Role of Life Cycle Inventory Databases for Life Cycle Assessment

The first LCA case studies consisted of about 50 processes, which were meant to reflect the entire life cycle (Gilgen et al. 1994; UBA 1995). This holds also for recent social LCA case studies (Ciroth and Franze 2011). These studies often took years to finish. In comparison, recent case studies contain hundreds to thousands of process data sets, and typically take less time and effort. This is only possible because of LCI databases available and in use for these case studies. Most of the processes are not generally modeled in the project but instead taken from the LCA databases.

Commonly, LCA studies are then distinguished into a foreground and a background system (Frischknecht 1998); the foreground system reflects the specific

---

<sup>3</sup>We implicitly define LCI data library here; in an LCI library, the provider of the datasets does not attempt to harmonize them, for example, to preserve the original modeling of the datasets. A “classic” example of an LCI library is ProBas, from the German Environmental Protection Agency, <http://probas.umweltbundesamt.de/>

product under study, while the background system is completed by using data from generic LCA databases.

This is more efficient than modeling common processes such as electricity and transport from the bottom up each time. It ensures consistency among practitioners that are performing LCA studies using the same database, and it makes realistic case studies in reasonable time possible. Often, the main contribution of the impacts in an LCA case study comes from the generic database, i.e., the background system, which is then more important than the foreground system for impact results. This shows the relevance of LCI databases for LCA and points at the importance of background datasets matching the goal and scope of the study.

On the other hand, since each database tries to provide one consistent, “safe modeling space,” this raises the question whether methodology and nomenclature of the database fit to methodology and nomenclature of the to-be-conducted study. Further, the unspecific, generic product provided in the database may not suit the specific product needed by the foreground system, which can be difficult for uncommon products (a specialty chemical, for example) or also for products from different regions (truck transport in India instead of truck transport in the European Union). Finally, different databases used in combination in one study may not fit together, as each of their “safe modeling spaces” might be inconsistent to each other, and the choice of one or the other database can have a strong influence on the overall result of a study. Some LCA studies performed a comparison between different LCI databases available for the building sector, where their methodology, documentation, data quality, and comprehensiveness were examined (Takano et al. 2014; Martínez-Rocamora et al. 2016). Based on their study, Takano et al. recommended enhanced information sharing between databases over developing newer databases. They also recommended the creation of a reporting and communication system for LCAs instead of trying to harmonize the methodologies among the databases.

Thus, nowadays professional, well-managed LCA databases seem essential for performing LCA case studies. They save time and effort and help to focus on specific, relevant aspects of the case study. On the other hand, the selected database can largely influence the modeled life cycle and calculated results in a case study. Therefore, providing and selecting a database for a study requires care.

### 3 Types of Databases

The first databases were created in the late 1980s and early 1990s. Meanwhile, more and more databases are appearing, and a market for LCA databases has been set up. A recent United Nations publication (Sonnemann et al. 2016, p. 56) lists about 40 different databases. As of today, after 3 years, only a handful databases have been updated, some new have emerged, and one major database was discontinued (Table 6.1).

Databases differ in various aspects. Some main aspects are mentioned here, with examples from Table 6.1:

**Table 6.1** Overview of selected LCA databases and libraries, country of origin, as well as number of datasets in 2012 and 2019

Name	Country (of database creator)	Type (DB: database, L: library)	Number of datasets as of 2012 (version or version year, if available)	Number of datasets, as of 2019	Comment
AGRIBALYSE	France	Process (DB)	822 (v1.2)	1188 (v1.3)	Last updated in 2015. AGRIBALYSE v3.0 is in development.
Agri-Footprint	Netherlands	Process (DB)	–	6342 (v4)	With some identical datasets from ELCD in different allocation models.
Australian Life cycle Inventory Database (AUsLCI)	Australia	Process (DB)	> 150	> 460	Database updates are ongoing.
Banco Nacional de Inventarios do Ciclo de Vida (SICV) <sup>§</sup>	Brazil	Process (DB)	10	22	Database updates are ongoing.
BioEnergieDat	Germany	Process (DB)	178	178	Last updated in 2012. There is no new project for updating the database.
Canadian Raw Materials Database (CRMD)	Canada	Process (DB)	13	18 (?)	
Chinese Life Cycle Database (CLCD v0.8)	China	Process (DB)	600	600	Last updated in 2011.
Ecobase	Chile	Process (DB)	147	147	The EcoBase database creation was a 2 year project that ended in 2015.
ecoinvent v3	Switzerland	Process (DB)	11,302 (v3.1)	16,024 (v3.5)	
ELCD 3.0 (European Life Cycle Database)	EU	Process (DB)	334	–	ELCD has been discontinued since 29 June 2019.
Extensions of ecoinvent data v.2.2	Switzerland	Process (DB)	6841	6841	Ecoinvent has since published ecoinvent v3 databases, the latest being ecoinvent v3.5, published in 2018.

(continued)

**Table 6.1** (continued)

Name	Country (of database creator)	Type (DB: database, L: library)	Number of datasets as of 2012 (version or version year, if available)	Number of datasets, as of 2019	Comment
GaBi LCA Databases 2019	Germany	Process (DB)	6513 (2013)	12,500 (2019)	Latest update available for 2019.
Inventory Database for environmental Analysis (IDEA) <sup>d</sup>	Japan	Process (DB)	3000 (v1)	3800(v2)	Last updated in 2016.
LCACommons <sup>i</sup>	USA	Process (DB)	–	9207	Subsuming the US LCI database.
LCADB.sudoe	Catalonia, Spain	Process (DB)	72	72	Last update unknown.
Mexicaniiuh <sup>h</sup>	Mexico	Process (DB)	81	81	No known update.
MY-ILCD <sup>b</sup>	Malaysia	Process (DB)	160	181	The Malaysian LCI database is not freely available, with only limited access to metadata level information.
NEEDS	International	Process (DB)	187	187	Not updated anymore.
Ökobau.dat (2014–2019-I) <sup>c</sup>	Germany	Process (DB)	954 (v2014)	1183 (v2019-I)	The current version of Ökobau.dat is 2019-I from 27 Feb 2019.
PEF tendered background datasets <sup>j</sup>	EU	Process (DB)	–	3504	Available for free for use within PEF, 2019.
ProBas	Germany	Process (L)	>8000	> 8000	Last known update in 2015.
Quantis Water Database <sup>f</sup>	Switzerland	Process (DB)	4000	4000	Last known update in 2012. Not publicly available database.
Quebec LCI database	Quebec, Canada	Process (DB)	900	900	The project for creating the database ended in December 2013.
SPINE@CPM <sup>a</sup>	Sweden	Process (DB)	>740	748	The first project was majorly funded by VINNOVA, Sweden's innovation agency between 1996 and 2006. It has been updated intermittently.

(continued)

**Table 6.1** (continued)

Name	Country (of database creator)	Type (DB: database, L: library)	Number of datasets as of 2012 (version or version year, if available)	Number of datasets, as of 2019	Comment
Thai National Life Cycle Database <sup>e</sup>	Thailand	Process (DB)	1300	1484	Last updated in 2017.
U.S. Life Cycle Inventory Database	USA	Process (DB)	880	880	Has become part of the LCACcommons.
eora <sup>k</sup>	Global	IO (DB)	14,839	14,839	Time series for the database available, 1990–2015.
exiobase <sup>l</sup>	Global	IO (DB)	9600 (v2.2)	9800 (v3.4)	Time series for the database available, 1995–2011.

<sup>a</sup> <http://cpmdatabase.cpm.chalmers.se/>

<sup>b</sup> <http://lcamalaysia.sirim.my/index.php/databases>

<sup>c</sup> <https://www.oekobaudat.de/en/database/database-oekobaudat.html>

<sup>d</sup> <http://idea-lca.com/?lang=en>

<sup>e</sup> [http://spaces.oneplanetnetwork.org/system/files/8b\\_database\\_roadmapping\\_key\\_considerations\\_thailand\\_11-17.pdf](http://spaces.oneplanetnetwork.org/system/files/8b_database_roadmapping_key_considerations_thailand_11-17.pdf)

<sup>f</sup> <https://quantis-intl.com/tools/databases/quantis-water-database/>

<sup>g</sup> <http://sicv.acv.ibict.br/Node/processList.xhtml?stock=IBICT>

<sup>h</sup> <http://www.centroacv.mx/mexicanuih.php>

<sup>i</sup> <http://www.lcaccommons.gov/>

<sup>j</sup> <http://eplca.jrc.ec.europa.eu/LCDN/contactListEF.xhtml>

<sup>k</sup> <http://www.worldmrio.com/>

<sup>l</sup> <http://www.exiobase.eu/>

- Databases may have a specific *regional scope*. They contain processes that represent a specific region. In part, this may be intentional or the result of practical limitations. For example, there are databases intentionally specific for one country (the Thai National database for Thailand, My-ILCD for Malaysia). There are databases intending to cover larger regions (ELCD, the European Reference life cycle database, for Europe), and databases with an intended global scope (ecoinvent, although originally started from Switzerland, with datasets from other regions being added over time; exiobase, with 49 countries and larger regions; eora, with overall 192 countries).
- Databases may have a *technical scope*, i.e., processes that represent specific technologies and provide certain products and services. As for the regional scope, there are databases that are intentionally broad and generic, and specific databases that focus on, for example, one industry sector. Quite a number of databases focus on agriculture (Agribalyse, Agrifootprint, ESU Worldfood), some on building components (ÖkobauDat). Other databases are intentionally generic/broad (GaBi professional, ecoinvent, exiobase, and eora). Similar to the regional coverage, the intended broad technological coverage can be more or less complete.

- Databases may differ in the *resolution of industrial activities* included. Roughly speaking, databases either report processes (“set of interrelated or interacting activities which transforms inputs into outputs,” ISO/TS 14048 (ISO/TS 14048: 2002<sup>4</sup>) following ISO 9000) or sectors (so-called I/O databases based on public statistics). In Table 6.1, eora and exiobase are I/O databases, whereas the other databases are process-based. For the process-based databases, some provide unit processes (again as defined in ISO/TS 14048, see also Chap. 3 in this book<sup>5</sup>), some only (or in addition) aggregated processes, with full or partial aggregation
- For using the database, also *organizational and procurement* aspects play a role. Some databases are free, some for purchase, with costs up to more than €10,000 per single-user license. Some databases are provided by public institutions (LCACommons, ELCD), some by private operators. Databases are furthermore updated at varying frequency, see Table 6.1.
- Databases further differ in their *quality assurance*. Most databases perform a review, some also by using external support from independent reviewers. These follow different review workflows and review schemes.
- Databases may differ in mere *technical aspects*, for example, in the implemented or supported import/export interfaces, the distribution “channel” of the database, for example, as part of an LCA software or stand-alone (see Sect. 4.4).
- Since databases aim to provide one coherent, consistent modeling space, it is evident that the *LCA methodology* may differ between databases, given that one universally accepted modeling approach does not exist yet. As a consequence, databases differ in various LCA choices, such as system boundaries, ways to deal with multifunctional processes and end of life, modeling biogenic, carbon and long-term emissions, to name just some of the typical LCA choices. A notable example are the three different “system model” databases provided by ecoinvent, which differ in the way they address end of life, allocation, and system expansion as well as linking processes.
- A related aspect is the supported *nomenclature* of the database, especially the supported elementary flow reference lists and supported LCIA methods.
- Finally, databases may differ in the addressed different *sustainability dimensions*. They may provide environmental inventory data, LCIA data, cost data, or also data about social impacts, alone or in combination, or also information only about climate-related impacts.

The first LCA databases were released in Switzerland (ecoinvent), Scandinavia (Sweden with SPINE@CPM, Finland with KCL-ECO), Germany (GaBi), Japan (IDEA), and in the US (USLCI), with a typically local data coverage. Over time, databases have been published also for other parts of the world, with focus on more

---

<sup>4</sup>ISO/TS 14048 (2002) Environmental management – Life cycle assessment – Data documentation format. 1st ed Geneva, Switzerland. ISO/TS 14048 was prepared by Technical Committee ISO/TC 207 “Technical Management,” Subcommittee SC 5 “Life Cycle Assessment” in 1993. This standard was last reviewed and confirmed in 2013, therefore the version of 2002 remains current

<sup>5</sup>Development of Unit Process Datasets



local processes and products (e.g., palm oil, (Archer et al. 2018), bananas, coffee, etc.), or different realizations of the same processes (truck transport in Brazil), often also linked to a capacity-building effort. One example is the recently concluded Sustainable Recycling Industries (SRI) initiative funded by the Swiss State Secretariat for Economic Affairs (SECO) in partnership with ecoinvent, where regional LCI networks (Brazil, Egypt, India, and South Africa) were set up to collaborate with the local networks for promoting capacity-building in developing LCI for the respective regions.<sup>6</sup>

The varying content of databases can be seen when plotting the number of datasets in a database per sector against the country. Using the same country names and the comprehensive UNSPSC<sup>7</sup> code for the sector classification, a plot of three databases that were all started with the idea to provide datasets as comprehensive and complete as possible shows major differences. Figure 6.2 shows an excerpt for the ELCD database, with countries in column A and UNSPSC sectors in line 2. Figure 6.3 shows the heatmaps for all compared databases, with dark background for better visibility. The ELCD database covers only a few sectors (mainly electricity, not readable of course from the plot) (a). The ecoinvent database v3.2 and 3.5 (band d) covers more sectors and has a focus on some countries, shown in the horizontal lines. The eora database has the most complete coverage of the three databases (c).

This short introduction shows the diversity of databases, which contrasts to the declared aim of database operators to create a “safe modeling space” for users, where datasets can be combined without major issues. The contrast comes evidently from that each single database that is in itself possibly consistent, but may not be

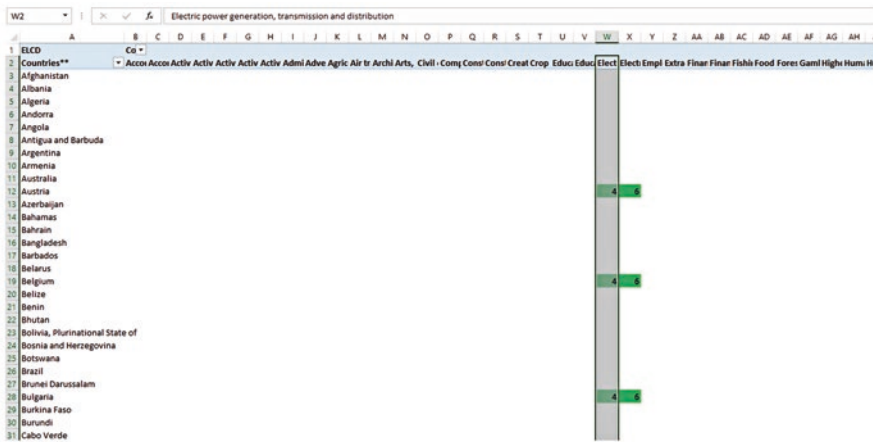
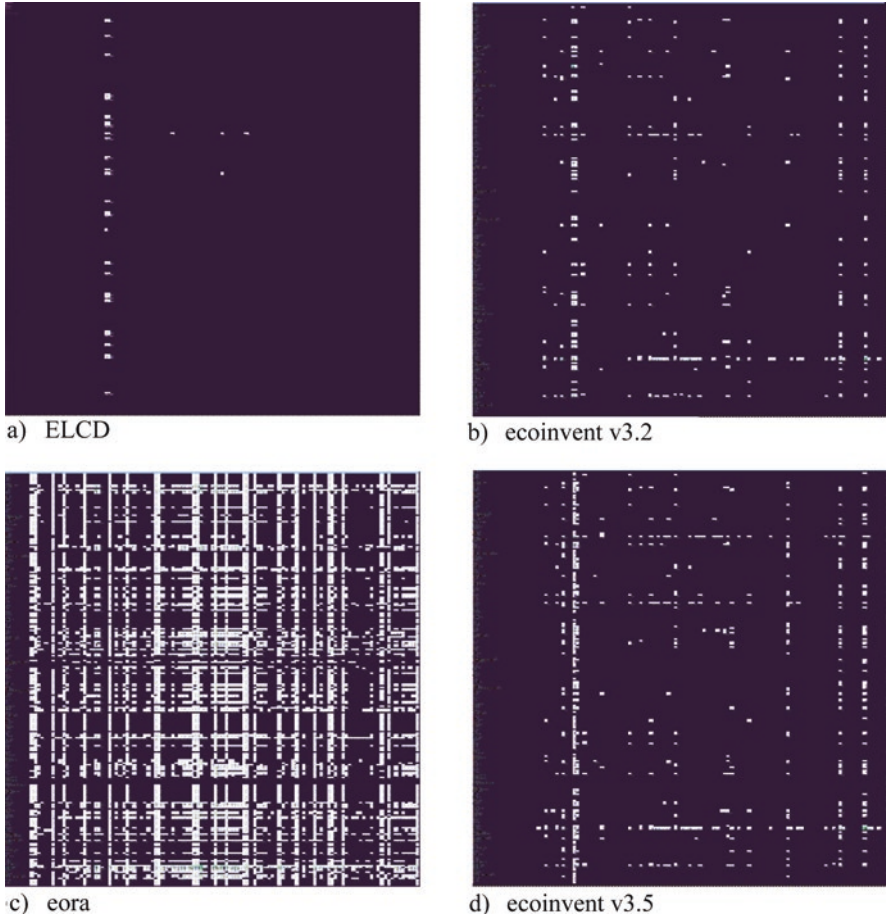


Fig. 6.2 Heatmap of the ELCD database, excerpt, countries, and sectors, with number of process datasets per sector in cells

<sup>6</sup> <https://www.ecoinvent.org/about/projects/sri-project/sri-project.html>

<sup>7</sup> The United Nations Standard Products and Services Code, <http://www.unspsc.org/>



**Fig. 6.3** Heatmap of different databases showing products and sectors (x axis) vs countries (y axis) covered

consistent compared to other databases; at the same time one single database may not be fully comprehensive so there is a need to reach across different databases. Therefore, for achieving a consistent space *across* different databases, three major approaches have emerged:

- The first approach is a network of consistent and aligned databases that promises to “expand” the harmonized data space and at the same time operate several databases independently. This idea was followed in the ILCD data network (JRC-IES 2010). Challenges in this solution are a harmonization of the datasets, i.e. to ensure that data sets are indeed aligned, as well as mere physical accessibility of the datasets, with sources of potentially varying reliability. A variant of

this approach is an integration of a separately created database into one larger database, where the separately created database follows the modeling of the larger database. This was performed when the Quebecois's database was integrated into the ecoinvent database (Lesage and Samson 2016)

- A second approach is a mild harmonization of databases, concerning flow nomenclature and LCA-modeling related aspects, and the provision of all these databases in one central “repository.” This is followed by the openLCA Nexus website<sup>8</sup> which is the largest repository of datasets available worldwide. An interactive map of the regionalized coverage of the datasets in the openLCA Nexus website is available in the Life Cycle Initiative website.<sup>9</sup> Obviously, a limitation is that processes in databases cannot be fully aligned; for system processes, mainly the nomenclature of flows can be changed, while the dataset modeling is “hidden” in the aggregation. Unit processes allow more changes, but an allocation applied to the process can be hardly changed, for example.

Both these options suffer from the limitation that they need to assume one specific modeling approach, flow nomenclature, and dataset use or set of uses, and try to apply this as consistently as possible. Possibly, the modeling approach is differentiated into several decision situations. For example, the ILCD handbook distinguishes decision support and accounting, and decision support with larger and small changes (ILCD Handbook 2010, pp. 38). This makes the database somewhat more flexible, but it still is unable to deal with many of the different modeling concepts and applications, which of course exist in “real-life” case studies.

Finally, as a further development, a system GLAD (see Chap. 5,<sup>10</sup> Sect. 3.5.3) was proposed and implemented in a first testing website (<https://www.globalca-dataaccess.org/>), through an international effort under the umbrella of UNEP/SETAC Life Cycle Initiative. The main idea is that several data providers submit datasets with “descriptors” that can be used to understand their modeling background and intended uses, which in turn allows users to specify what they are interested in, and find datasets that best match their needs. Section 6 in this chapter, the Outlook, spends some thoughts on this concept and its further development.

## 4 Issues in Life Cycle Inventory Databases

Creating and maintaining an LCI database presents issues and challenges in several aspects, including setup, maintenance, finances, quality assurance, and not the least integration into LCA software.

---

<sup>8</sup><https://nexus.openlca.org/>

<sup>9</sup><https://www.lifecycleinitiative.org/applying-lca/lca-databases-map/>

<sup>10</sup>Data Quality in Life Cycle Inventories

## 4.1 Setup

Setup here means the starting phase of a database. Since a database aims to provide a “safe modeling space” (see definitions in Sect. 1), the initial questions to consider are:

- (i) Which datasets should be provided in the database?
- (ii) In which “sequence” should they be created and provided?
- (iii) Which modeling conventions should be followed by the database?

Also, of course, the following questions should be clarified at the setup phase:

Of the technical solution used  
Of longer-term maintenance  
Of financial and operational sustainability  
Of quality assurance

The decisions taken at the setup phase determine the scope of the database, and eventually, ensure the success and long-term usability of the database. Current and previously existing databases may have taken different decisions for the setup or they have been influenced by their operators and initiators. This is evident in the varieties of LCA databases available, see Sect. 3.

Regarding the database content, i.e., the processes in the database, all databases need to solve the issue of where to start with modeling and which becomes delicate regarding closed loops existing in production systems (UNEP/SETAC 2011). For example, production of steel needs steel used in processing machinery and the production of diesel needs diesel for transport. This self-reference of LCI database systems is evidently more complicated for databases that contain aggregated processes than for databases that contain unit processes, since unit processes can be modeled also without access to, and knowledge of, the full life cycle chain. However, thinking of the consistent modeling space a database aims to provide, also a unit process database initially needs to complete supply chains with links to datasets from other databases, or already include these datasets, which in both cases raises the question of how well the other database fits to the own modeling.

Discussed here are the strategies for the setup in order to provide datasets comprehensively. For the one-sector databases (see Sect. 3), the situation is comparable, with the limitation that they never, by intention, will be able to fully provide complete life cycles.

*First*, a database can follow a bootstrapping approach, by starting from those processes that are most commonly used and needed by other datasets and by users. Often, these are transport and electricity, followed by construction and basic materials.

To take just one example, for “rubber sandals and slippers” from the Japanese IDEA database, the overall product system contains about 1600 individual processes, but some of them are used very often in the product system. The top five

**Table 6.2** Top five most often used processes in a product system created for the Japanese IDEA database, for the product “rubber sandals and slippers” (taken from openLCA 1.8)

Processes	Number of linked inputs
331111014 electricity, Japan, 2014FY	1278
361111000 tap water	1138
181114801 energy, kerosene combustion	1094
341111801 energy, town gas 13A combustion	1092
181124801 energy, liquefied petroleum gas (LPG) combustion	1063

most used are electricity, tap water, kerosene, town gas, and liquefied petroleum gas combustion (Table 6.2).

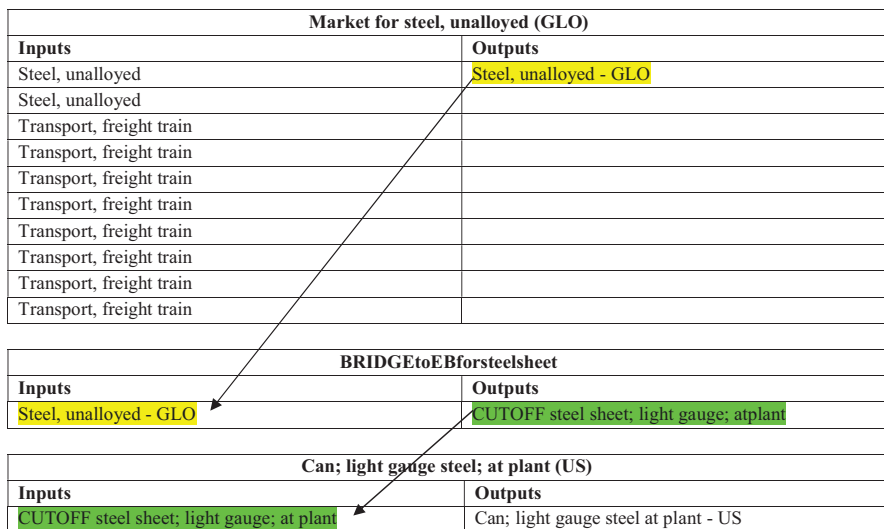
If these datasets are initially created, they can be used many times in the product system; and if the database development focuses on those datasets that are used most often across the targeted overall datasets, the database can ideally grow, building on datasets that have already been created. This approach has been used by the Chinese CLCD database (Wang et al. 2011), where first versions contained transport and energy datasets only, with construction datasets being added in later versions.

*Second*, a database can start by creating datasets for several sectors and independent products in sub-projects at the same time, and share only aggregated datasets, in a limited extent, between these sub-projects. This evidently risks that datasets might become inconsistent, if several projects use differing datasets for, for example, electricity in their supply chain. Motivations for this approach might be capacity restrictions, time pressure, and the desire to involve several parties in the creation of the database early on. This approach has been used in the creation of the “EF-compliant” datasets, where about 12 different tenders have been launched to create parts of an Environmental Footprint background database.<sup>11</sup> Those tenders were awarded to different consultancies as well as institutes, and started with only little overlap. As a consequence, the datasets of the first of these tenders, for energy and transport, were available only for the very late data tenders, and most of the datasets could not be shared across the projects.

*Third*, a database can let supply chains intentionally open, in that the database creator does not attempt to provide all process datasets needed to deliver all required products, but keeps links open. Ingwersen et al. (2018) propose to rename these unfollowed products “CUTOFF,” and to additionally provide “bridge processes” in the database that link them to specific background databases (Fig. 6.4).

*Fourth*, and final, it is possible to include datasets from other databases to complete supply chains. The license of the other database needs to permit this, and also the modeling approach should be somewhat aligned, which is often challenging. For example, the Agri-Footprint database completes agricultural dataset supply chains with aggregated datasets from ELCD (European Reference life cycle database).

<sup>11</sup> [http://ec.europa.eu/environment/eussd/smgp/ef\\_pilots.htm#compliant](http://ec.europa.eu/environment/eussd/smgp/ef_pilots.htm#compliant), accessed April 21, 2019



**Fig. 6.4** Bridge processes and cutoff flows to make a database more flexible and to preempt a database from providing all products used (Ingwersen et al. 2018)

Whereas it provides its own datasets in different allocation models (price, energy, mass), the ELCD aggregated datasets cannot be changed.

Independent from the scope of the datasets and supply chains included in the database, setting up the database should also determine and set up the technical infrastructure for the database, with review workflow procedure, a tool for entering data and for moderating updates, physical databases to store the information, release channels for data, and not the least appropriate communication channels and measures.

## 4.2 Quality Assurance

Quality assurance is an essential part of a database, right from its creation to its long-term maintenance. A sound quality management includes the following points:

- *Goal and scope for the datasets* in a database should be clearly specified so that data quality can be determined (see Chap. 5 in this book). Ideally, the specification should cover reference time, location, the products to be modeled, and also LCA modeling issues such as system boundaries, dealing with multifunctional processes, modeling of waste, water, biogenic carbon, and long-term emissions
- *A system for assessing the quality of datasets* is in place, meaning that the data quality criteria and their assessment are specified and documented, and that a procedure and infrastructure is implemented to allow an execution of the assessment. The infrastructure includes technical tools, as well as accredited or recognized experts who can perform the review

– Conformance to the *data quality topics* mentioned in ISO 14044 is a good starting point for creating the quality guidelines for assessing the data quality standard of a database. These topics can be summarized into four key areas:

- Representation and conformance aspects (time, geography, technology)
- Modeling-related aspects (selected nomenclature, modeling waste, biogenic carbon, multifunctionality)
- Measurement-related aspects (completeness, reliability of the source, uncertainty of data)
- Procedural aspects (review procedure, copyright)

Figure 6.5 shows the review procedure followed by ecoinvent for validating a unit process. The procedure involves three different actors, including an ecoinvent manager that prepares the dataset for the database, followed by the due-diligence process carried out by two ecoinvent experts where the dataset is checked for significant issues, completeness, mathematical correctness, plausibility checks, sensitivity, uncertainty, and consistency on the basis of their quality guidelines. Corrections or modifications wherever necessary are carried out prior to the creation and documentation process. The whole process is a reiterative and takes place parallel to another until a satisfying dataset is achieved.

Datasets in a database have to undergo and pass quality assurance. If the data quality assessment includes more than a binary passed/not passed result, the assessment result, and in all cases comments regarding quality assurance of the datasets, should be provided along with the datasets.

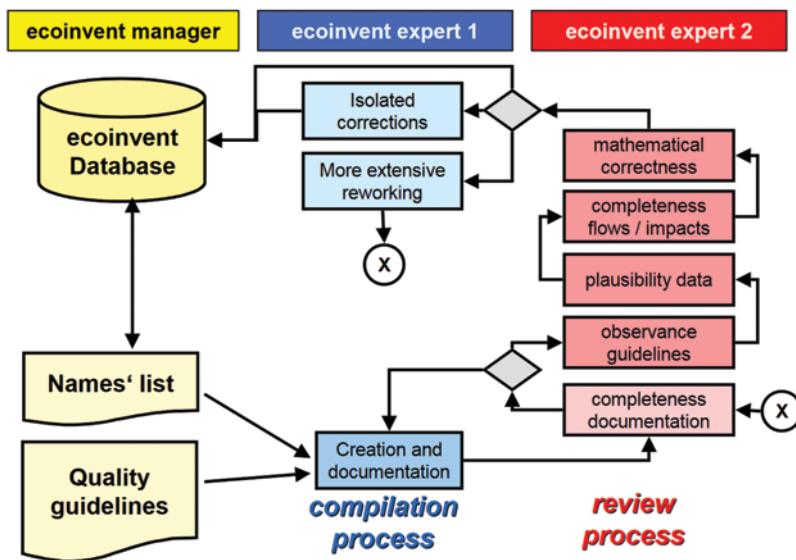


Fig. 6.5 Overview of the internal review procedure within the ecoinvent database (Frischknecht and Jungbluth 2007, p 54)

The quality of a database and of datasets in the database is crucial for the success of the database on the market. Database providers typically emphasize the performed quality assurance and review; some even provide documents about *external* quality assurance. Several frameworks exist today to address “data quality” in a database; the few notable ones are developed by:<sup>12</sup>

- US EPA (United States Environmental Protection Agency)
- UNEP/SETAC Life Cycle Initiative (United Nations Environment Programme, and Society of Environmental Toxicology and Chemistry)
- European Commission

For users of a database, a performed quality assurance is reassuring in that the database fulfills its purpose to provide a “safe modeling space.” The detailed quality indicator results can, however, hardly be checked by a database user. For example, it is almost impossible to trace back whether a dataset refers to 2012 or 2014. This makes a clear documentation of datasets and data sources used even more important; and even more so, a documentation of the deviation from the intended goal and scope set forward for the entire database, be it intended deviation, or a deviation rather done as concession to practical requirements. This “helplessness” for a user to verify a database modeling is especially challenging for aggregated datasets where the underlying detailed model cannot be accessed by the user.

### 4.3 Maintenance

Maintenance of an LCI database means the provision of updates to datasets to reflect technical changes in the real world (lower emissions of cars, more efficient electricity generation, to name just two), and also the update of the database content to align with progress in LCA and especially LCIA. Newer LCIA methods often create a need for more detailed elementary flow sets (from “dust” to “PM 2.5”, for example, see Chap. 9 in this book),<sup>13</sup> and it is typically expected that a database is expanded, i.e., contains an increasing number of datasets over time.

Maintenance is often a survival issue for databases. There is a long list of databases that disappeared after some time, despite having been created with initially enough funding and resources. It is commonly stated that database maintenance is important (UNEP/SETAC 2011), but when a database project is initiated, it is typically unclear what its long-term future looks like.<sup>14</sup> Funding for a continuation might typically be in sight only after a successful first project.

---

<sup>12</sup>For US EPA, see “Data Quality”, Chap. 5, Sect. 3.5.2, for UNEP/SETAC, see GLAD, Global Life Cycle Access to Data, Chap. 5, Sect. 3.5.3; for European Commission, see Environmental Footprint, Chap. 5, Sect. 3.5.1)

<sup>13</sup>The link between life cycle inventory analysis and life cycle impact assessment

<sup>14</sup>Since no database starts with guaranteed financial support over an unlimited time



While predicting the future for one given database project is difficult, it is easier to identify elements that contribute to a longer-term existence and maintenance of the database. These are:

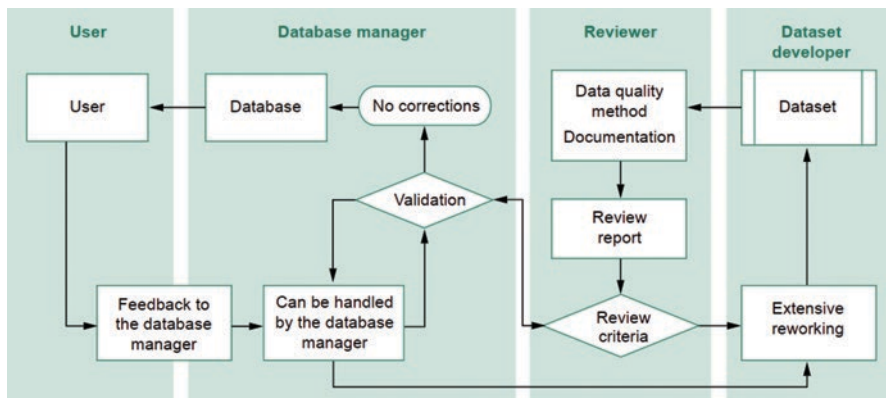
- Financial support from public sources
- License fees
- Relevance: A sufficiently sized database or relevant dataset that is requested and not available elsewhere, which are recent
- Trust: An established name, maintained with quality assurance, documentation, and communication
- Ease of access and availability

The financial aspect is not to be underestimated. In the end, an LCI database is a product that requires considerable initial effort and long-term resources for the maintenance, and thus needs a sound business plan. In particular, a new database needs to compete with other LCA databases on the market, which also requires effort.

*Public support* plays a significant role among the elements for database sustainability. For one, it helps to lower license fees and to provide a database that does not meet market demand initially. On the other hand, one could argue that sustainability data are common good and therefore *should* be provided for free, with public support covering the expenses, just as street lights or other infrastructure (De Rosa et al. 2017). Public support likely makes a database more prominent (“this is the database supported by the European Commission”), but it can also make progress slower and more bureaucratic, and political changes can lead to rather abrupt changes in database development, even to discontinuation. Public support consists often of only one or very few supporters, and a change in the organizational structure can put these few sources at risk.

*License fee* income, on the other hand, can help to focus on market needs, and is certainly a more broadly spread and stable source of income once the database is established. Reaching this level of establishment is, however, challenging, since with initially low license fee incomes, a lot of work needs to be spent on dataset creation and on establishing the required infrastructure, and in addition, a database typically competes with other existing databases on the market.

Financial aspects aside, it is always in the interest of a database to be used. Maintenance is performed also to keep and extend the user base of a database, and to keep the database relevant. Adding and updating datasets is one core aspect of maintenance. Typically, databases follow a dedicated workflow for adding and updating datasets, with several actors involved: dataset developers, reviewers, the database managing team, and users. Figure 6.6 shows a possible workflow. Dataset developers create a dataset, send it to a reviewer, who checks the dataset following a database-wide data quality approach, writes a report, sends it (probably condensed to review criteria results) back to the developer, or to the database manager, who checks and validates it in terms of whether the dataset represents what it is intended to represent. There might be iteration loops between this validation, the review process and review criteria assessment. If the dataset is found sufficient, it is integrated into the database, possibly first in a staging version of the database. Otherwise, the



**Fig. 6.6** Overview of a database management structure with focus on dataset creation and update (UNEP/SETAC 2011, p. 94)

dataset is improved by the database management team or by the dataset developer. Finally, database users provide feedback to the database management team, which is hopefully considered.

A database can be extended and updated with new large projects, in a more organic way using license fees. It can also be updated by third parties that provide and contribute to the datasets, thereby using the database as a publication platform and benefit from the review procedure including quality assurance of the database.

#### 4.4 Integration into LCA Software

Technical accessibility of a database typically helps increasing the user base. In the end, a database will not be used primarily stand-alone, but to calculate and understand life-cycle impacts, which requires calculation software. It is therefore in the interest of a database to collaborate with software providers, and to ease the integration of the database into LCA software. Therefore, all major databases are now partnering with LCA software providers, either to establish contractual relationships that enable software companies to resell the databases, withholding a reseller rebate and thus creating an incentive for the software provider, or to provide the database in one of the common LCA data exchange formats to allow easy import of the database by the software users directly.

Only very few databases are created, supported, and provided by an LCA software developing company.<sup>15</sup> On the other hand, only very few databases are published without being either integrated in the software or available in an exchange

<sup>15</sup>For example, Thinkstep, with the GaBi LCA software and the various GaBi databases, and GreenDelta, with the PSILCA social LCA database.

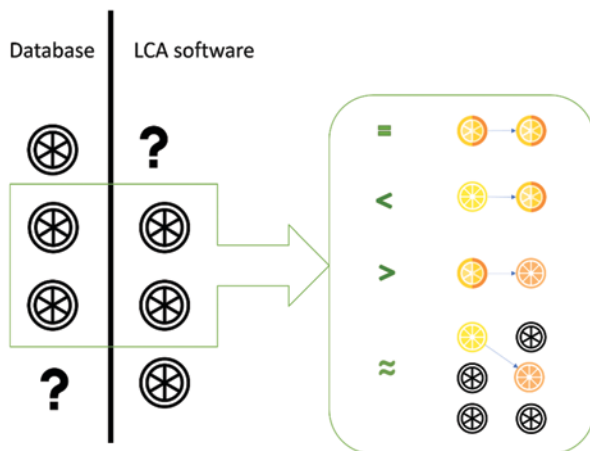
format; one exception was the FEFCO database of the European Paper and Cardboard association that was provided on a printed brochure only until about 2015.

While the integration of databases in LCA software is often essential to reach users, it requires adapting the database to the structure of the software in two main ways:

1. Information contained in the database needs to be mapped to the available fields in the LCA software database and user interface. Information that is not considered in the software can either be put in comment fields or omitted. Two examples are provided:
  - The ecoinvent database considers exchange properties that describe water and carbon content and other properties, of every exchange. For example, for an emission of a tin ion, the carbon and water content are reported. These properties are not included in any LCA software so far.
  - The datasets tendered by the European Commission for the Environmental Footprint changed the way to model locations in 2018. Previously, one flow had one location assigned (emission of ammonia to air in the Netherlands). Now, a so-called exchange, i.e., a flow that is input or output of a process, has the location assigned (emission of ammonia to air in the Netherlands, from animal husbandry); this avoids the creation of thousands of flows for all different locations, but requires that the impact assessment calculation considers the exchanges instead of the flows, which is not supported so far by any of the major LCA software packages, and it requires of course that the software can store and show in the user interface the location of the exchanges.
2. The database flow nomenclature needs to be adapted often to fit the nomenclature and categories of the software. There, several different compatibility situations can occur: an exact match (=), less generic to more generic (<), more generic to less generic (>), and proxy (~).

Figure 6.7 illustrates cases for database integration in an LCA software. Suppose a database is to be mapped to the software, with its internal database structure and software-specific LCA reference data. The black circular icons represent elements that are present in the respective structures. The question mark across the element of either the database or the software indicates that the particular element is missing in one of them. Where there are elements matching between the database and the software, the match could be in either four ways. Considering example flows (or reference flows) of citrus fruits, match 1, “=,” is a full match, for example, when the flow is citrus fruits in both databases. Match 2 and 3 (“<” and “>”) indicate more generic mappings, where the more generic flow (e.g., fruit) is in case 2 on the software side and in case 3 on the database side. Case 4, finally, is a proxy mapping, “~”; in the example, a citrus flow might be mapped to a flow representing an orange. This simple example is valid for any of the information in the database, and for any information considered by the software.

A combination of different LCA databases into one software raises additional considerations. Taking one database as the attempt to provide one harmonized,



**Fig. 6.7** Database integration in an LCA software, mapping cases. For further explanation, see Sect. 4.4

consistent modeling space calls for a user- or software-provided strategy to deal with data from different model perspectives and concepts in one software.

## 5 Data Exchange

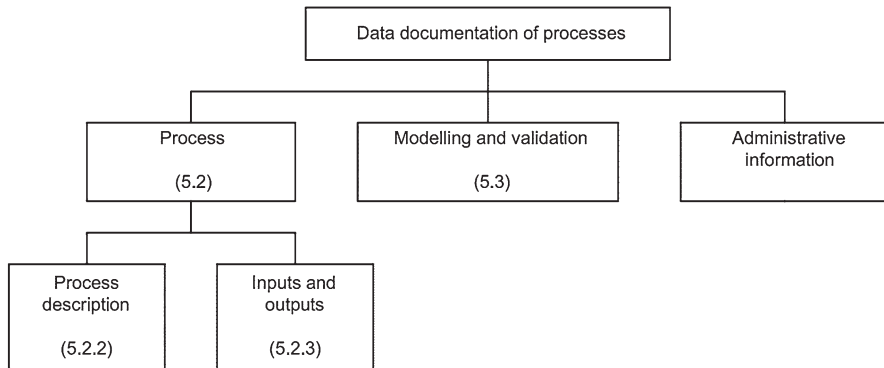
### 5.1 Information in LCI to Be Exchanged

As shown in Fig. 6.1, all the information in an LCA allows data exchange.

For a process dataset, which is often exchanged, Fig. 6.8 presents further details, as proposed by ISO/TS 14048 (ISO/TS 14048: 2002), and adds administrative information, such as dataset owner, dataset creation date, among others, as well as documentation of modeling and quality assurance. Box 5.2.3 in the figure represents the exchange data that primarily contains the input and output flows, their respective direction, amount, and flow property, and is further supported by data quality information, including parameters or dataset-specific formulae, among others. Further, administrative and modeling information can be provided for flows, for product systems, and LCIA methods, for example.

### 5.2 Exchange Formats

Some sort of exchange formats for LCA data existed as early as around 1980 when the first LCI and LCA databases appeared. The release of the ISO/TS 14048 standard attempted to align the different concepts, by proposing a data documentation



**Fig. 6.8** Data documentation elements for a process dataset, as proposed by ISO/TS 14048

format for processes as shown in Fig. 6.6. Since then, all newly released formats for LCA data refer to ISO 14048/TS and can be considered compliant.

A software-integrated database differs technically on the basis of its linking concepts, size, documentation fields, field separators, to name a few, and is probably not directly accessible without the LCA software. Most databases have to be further modified and adapted to different LCA software. Typically, a database is designed for a specific software; however, it is not user-friendly to switch software for including different databases. An exchange format promises to contain “the important” information and allow an exchange from one user to another, and even from one software to another, without considering the software-internal database structure.

Over time, also the exchange formats for LCA have evolved. Nowadays, four formats are frequently used, see the following. Three of the four formats are XML formats, i.e., they follow the extensible markup language.<sup>16</sup> One format follows JSON-LD, Java Script Object Notation for Linked Data.<sup>17</sup>

*EcoSpold 1*<sup>18</sup> is the format initially released with the ecoinvent2 database, created for the ecoinvent center. It is the oldest of the ISO 14048 compliant exchange formats that is still in use. The file format goes back to an association created in the 1990s by a group of companies and researchers, forming the Society for the

<sup>16</sup> XML was first proposed by the W3C consortium in 1998 with the idea to provide a language that is usable over the internet, easy to write and to process, with formal and concise design, among other things (<https://www.w3.org/TR/1998/REC-xml-19980210>). It has been broadly adopted since.

<sup>17</sup> JSON-LD was developed with support from search engines, to overcome some disadvantages of XML in data interchange; it is now a recommendation of W3C (<https://www.w3.org/TR/json-ld/>). In JSON, data is not organized in a hierarchical tree as in XML but in a “map” (<https://www.educba.com/json-vs-xml/>), with simple annotations, with makes processing faster and the overall format less heavy. JSON-LD is simply speaking JSON for linked data, so that the format can directly represent ontologies for example.

<sup>18</sup> [https://www.ecoinvent.org/files/ecospold1\\_data\\_exchange\\_format.zip](https://www.ecoinvent.org/files/ecospold1_data_exchange_format.zip)

Promotion of Life-Cycle Data (SPOLD) with the aim to create one common format for LCA. This initiative led to the creation of the first SPOLD data format.<sup>19</sup> It is relatively easy, cannot distinguish processes from products, and does not understand parameters. The format is supported by almost all existing LCA software systems. Result is typically one single XML file for one process.

*EcoSpold02*<sup>20</sup> is the format developed on behalf of theecoinvent center for the ecoinvent 3 database. This format understands parameters, unique identifiers, and distinguishes processes, flows, units, and other elements. It furthermore has many different, detailed features, for example, properties for exchanges. Being also an XML format, it has so far only been implemented in the openLCA software, apart from ecoinvent's own dataset editing software ecoEditor that is not intended for LCA calculation.

*ILCD*<sup>21</sup> is the format developed for the Joint Research Center (JRC) as reaction of some shortcomings of the EcoSpold1 format. It was released before EcoSpold02, and the first supporting database was ELCD. Being an XML-based format, it understands unique identifiers, parameters, and distinguishes processes, flows, unit groups, and other elements. All these elements are provided in one folder structure, as single XML files, and overall as one zip archive. Several extension formats exist meanwhile, for example, the ILCD+EPD format to specifically address environmental product declarations.

*JSON-LD*,<sup>22</sup> finally, is the newest of the formats, developed in JSON-LD, for openLCA. It supports parameters and unique identifiers. Similar to the ILCD format, it distinguishes different elements for an LCA model (product system processes, flows, etc.), which are stored in a folder structure and can be exchanged as a zip archive. Due to the more efficient information storage, datasets need roughly 50% of the space of the ILCD format; the datasets link directly to semantic web and ontology spaces.

Overall, data formats are quite different in the way they store information (file format) and also in details, but they all cover the majority of information to be exchanged. However, not all formats have mandatory fields to be considered by other formats.

One important aspect is that the exchange format is not necessarily identical to the format in which a database stores information. Rather, it is literally meant for *exchanging* information. Databases can thus be designed to support several data formats.

---

<sup>19</sup>Weidema B. SPOLD '99 format – an electronic data format for exchange of LCI data (1999.06.24) <https://lca-net.com/files/sis.pdf>

<sup>20</sup>[https://www.ecoinvent.org/files/documentation\\_on\\_ecospold2\\_format.v1.0.13.zip](https://www.ecoinvent.org/files/documentation_on_ecospold2_format.v1.0.13.zip)

<sup>21</sup><https://eplca.jrc.ec.europa.eu/LCDN/developerILCDDDataFormat.xhtml>

<sup>22</sup><http://greendelta.github.io/olca-schema/>

### 5.3 Interoperability Concepts

Formats with differing ways to present information, and LCA users relying not only on one single, coherent modeling space but on different approaches (e.g., due to regional conventions and innovations), created the need to address interoperability in LCA data and databases. Consequently, several elements have been developed to meet this demand; they comprise:

- A format converter to convert between different LCA exchange formats.<sup>23</sup> An alternative approach is to use the import and export features of LCA software. This was done in the GLAD server where the openLCA software with its import and export interfaces is integrated to enable data format conversion.<sup>24</sup>
- A better alignment of data formats, to prevent clashes between data formats, where mandatory fields in one data format have no corresponding field in another one.
- Mapping files to align categories and nomenclature for flows and other elements.<sup>25</sup>
- A deeper understanding and possible conversion of modeling-related aspects, which is, to some extent, the aim of the GLAD system (see Chap. 5, this book<sup>26</sup>).

It seems fair to say that these elements, despite being useful, at present do not fully permit a fluent switch from one database to another, or a seamless combination of different databases. One reason is certainly that a seamless combination of different databases somewhat contradicts the original idea of a database as one harmonized modeling space; increasing diversity in databases and increasing user demand might, in future, indeed allow this combination and enable this shift.

## 6 Outlook

Databases constitute a foundation for today's large, comprehensive LCA case studies, and yet, creation and maintenance require considerable effort, and exchange across different databases is not fully solved today. We expect that smarter ways for collecting data become increasingly important to make data collection faster, less error-prone, and easier. It can also be expected that data exchange, also across different software systems and different modeling choices, becomes standard, and that eventually the data material collected in databases will gain in more comprehensiveness and topicality, with possibly event-based information to be added. Novel IT developments can play a role. The JSON-LD format might replace today's prevalent

---

<sup>23</sup><https://github.com/GreenDelta/olca-converter>

<sup>24</sup><https://github.com/GreenDelta/olca-conversion-service>

<sup>25</sup> <https://github.com/USEPA/Federal-LCA-Commons-Elementary-Flow-List/tree/master/fedelemflowlist>

<sup>26</sup>Data quality in life cycle inventories

XML for data exchange formats, and blockchain approaches might be used for documenting supply chain interactions, see, for example, Kim and Laskowski (2018). Still, providing and maintaining an interoperable, relevant database will probably always be challenging, and there is probably no easy “silver bullet” technical solution. It seems hard to believe that one single technology, be it blockchain or other, will be able to provide a perfect solution; rather, a balanced portfolio of new and established technologies as well as procedures seem to have the potential to indeed change the way databases for life cycle inventories will be used in future, hopefully leading to more reliable, comprehensive, interoperable, and relevant LCI databases.

## References

- Archer SA, Murphy RJ, Steinberger-Wilckens R (2018) Methodological analysis of palm oil bio-diesel life cycle studies. *Renew Sust Energ Rev* 94:694–704
- Ciroth A, Franze J (2011) LCA of an Ecolabeled notebook: consideration of social and environmental impacts along the entire life cycle. [greendeltatc.com](http://greendeltatc.com)
- De Rosa M, Weidema B, Merciai S, Schmidt J (2017) LCM Luxemburg, September 6 th 2017. Making the dream come true: sharing linked data and software with BONSAI
- Frischknecht R (1998) Life cycle inventory analysis for decision-making: scope-dependent inventory system models and context-specific joint product allocation. ESU-services, Uster. Research-collection ETHZ Switzerland
- Frischknecht R, Jungbluth N (eds) (2007) ecoinvent 2 overview and methodology. *ecoinvent Rep no 1* 1:1–77
- Gilgen P-W, Ackermann R, Fleischer G (1994) In: Fleischer, G. (ed.) *Stoff-und Energieflussanalyse für die Aluminium-Dose. Produktionsintegrierter Umweltschutz*, EF, Berlin, ISBN: 978-3-924511-81-4
- ILCD Handbook (2010) Joint research general guide for life cycle assessment – Detailed Guidance Centre. Institute for Environment and Sustainability, Ispra
- Ingwersen WW, Kahn E, Cooper J (2018) Bridge processes: a solution for LCI datasets independent of background databases. *Int J Life Cycle Assess* 23:2266–2270
- ISO/TS 14048 (2002) Environmental management – life cycle assessment – data documentation format, 1st edn. ISO, Geneva
- JRC-IES (2010) Specific guide for life cycle inventory data sets. Joint Research Center, Ispra
- Kim HM, Laskowski M (2018) Toward an ontology-driven blockchain design for supply-chain provenance. *Intell Syst Account Finance Manag* 25:18–27
- Lesage P, Samson R (2016) The Quebec life cycle inventory database project. *Int J Life Cycle Assess* 21:1282–1289
- Martínez-Rocamora A, Solís-Guzmán J, Marrero M (2016) LCA databases focused on construction materials: a review. *Renew Sust Energ Rev* 58:565–573
- Sonnemann G, Strothmann P, Weyand S, Valdivia S (2016) Opportunities for national life cycle network creation and expansion around the world. Technical report. <https://www.researchgate.net>
- Srocka M (2009) openLCA: open source software für life cycle assessments – Stand und Weiterentwicklung. In: Feifel S, Walk W, Wursthorn S, Schebek L (eds) *Ökobilanzwerkstatt 2009*. KIT Scientific Publishing, Karlsruhe, pp 11–20
- Takano A, Winter S, Hughes M, Linkosalmi L (2014) Comparison of life cycle assessment databases: a case study on building assessment. *Build Environ* 79:20–30
- UBA (1995) Schmitz S, Oels H-J, Tiedemann A: *Ökobilanz für Getränkeverpackungen, Teil B: Vergleichende Untersuchung der durch Verpackungssysteme für Frischmilch und Bier hervorgerufenen Umweltbeeinflussungen*. UBA Texte 52/95. Berlin 1995



- UNEP/SETAC (2011) Global guidance principles for life cycle assessment databases – “Shonan Guidance Principles.” UNEP/SETAC
- Wang H, Wang Z, Fan C et al (2011) Development of Chinese reference life cycle database (CLCD) – guidelines, documentation and tools. Life Cycle Manag Conf LCM