LCA database creation:

Current challenges and the way forward

Andreas Ciroth[a*], Claudia Di Noi[a], Salwa Syed Burhan[a], Michael Srocka[a]

[a]GreenDelta GmbH, Kaiserdamm 13, 14057 Berlin, Germany

`

First submission: 21st July 2019, Revised submission : 21st November 2019, Acceptance: 23rd December 2019

Abstract

LCA studies require a high volume of data and their quality has a direct influence on the quality of the Life Cycle Inventory (LCI) and Life Cycle Assessment (LCA) study overall. The use of LCA databases enables users to (i) reduce time, efforts, and resources for data collection and (ii) reflect supply chains they have no direct control over. On the other side, it creates the need to align own modeling of the foreground LCA study with the modeling in the database. In recent years, countries worldwide have been more and more motivated in supporting LCA studies by providing national databases that reflect their economy, energy mix, and disposal technologies. This article aims to give insights on the main needs, requirements, and challenges for the creation of an LCA database, with a special focus on national, reference databases. First, the article defines the main characteristics of LCA datasets and discusses data collection approaches. Secondly, LCA databases are defined, and the creation of LCA databases from developed datasets is addressed, including the case of national LCA databases. Finally, the existence of tools that could ease the LCA dataset and database creation process is investigated, namely the LCA Collaboration Server and the LCA Data-Machine. It is important that countries willing to create a national database are supported, for example with capacity-building workshops, by actors with a long tradition in the field, which is of mutual benefit: Countries with a long tradition in LCA will benefit from interactions with newcomers, for instance by discussing together unsolved methodological and interoperability issues; newcomers do not need to start from scratch but can benefit from gained experiences. Creating databases that provide specific data for various parts of the world supports LCA methodology and application in general, and it is not the least a chance for local LCA communities to bring in innovation into LCA, and benefit from existing experiences at the same time.

Keywords: *database; dataset; data; LCI; tool.*

Abstrak

Studi LCA membutuhkan volume data yang tinggi dan kualitasnya berpengaruh langsung terhadap kualitas kajian Life Cycle Inventory (LCI) dan Life Cycle Assessment (LCA) secara keseluruhan. Penggunaan basis data LCA memungkinkan pengguna untuk (i) mengurangi waktu, upaya, dan sumber daya dalam pengumpulan data dan (ii) mencerminkan rantai pasokan yang tidak dapat mereka kendalikan secara langsung. Di sisi lain, hal ini menciptakan kebutuhan untuk menyelaraskan pemodelan pada studi LCA yang dibuat sebagai *foreground* dengan pemodelan dalam database. Dalam beberapa tahun terakhir, negara-negara di seluruh dunia semakin termotivasi dalam mendukung studi LCA dengan menyediakan database nasional yang mencerminkan ekonomi, bauran energi, dan teknologi pembuangan mereka. Artikel ini bertujuan untuk memberikan wawasan tentang kebutuhan, persyaratan, dan tantangan utama dalam pembuatan database LCA, dengan fokus khusus pada database referensi nasional. Pertama, makalah ini mendefinisikan karakteristik utama kumpulan data LCA dan membahas pendekatan pengumpulan data. Kedua, makalah ini akan mendefinisikan basis data LCA, dan menelaah pembuatan basis data LCA dari kumpulan data yang dikembangkan, termasuk kasus basis data LCA nasional. Terakhir, studi ini mendalami keberadaan alat yang dapat memudahkan dataset LCA dan proses pembuatan database, yaitu LCA Collaboration Server dan LCA Data-Machine. Negara-negara yang ingin membuat database nasional didukung dan dukungan ini menjadi penting, misalnya dengan lokakarya peningkatan kapasitas, oleh para pelaku dengan tradisi panjang di lapangan, yang saling menguntungkan: Negara-negara dengan tradisi panjang di LCA akan mendapat manfaat dari interaksi dengan pendatang baru, misalnya dengan mendiskusikan bersama-sama masalah metodologis dan interoperabilitas yang belum terpecahkan; pendatang baru tidak perlu memulai dari awal tetapi bisa mendapatkan keuntungan dari pengalaman yang didapat. Pembuatan database yang menyediakan data spesifik untuk berbagai belahan dunia akan mendukung metodologi dan aplikasi LCA secara umum, dan ini merupakan kesempatan bagi komunitas LCA lokal untuk membawa inovasi ke dalam LCA, dan sekaligus mendapatkan manfaat dari pengalaman yang ada.

Kata Kunci: *basis data, kumpulan data, data, LCI, alat*

* Corresponding author. Tel.: +49 30 914 744 29; Fax: +49 30 484 969 91
ciroth@greendelta.com

1. BACKGROUND AND MOTIVATION

Life Cycle Inventory (LCI) is recognized as the second phase of a Life Cycle Assessment (LCA) study [1,2]. LCI is often considered the most time- and resource-consuming step in LCA as it consists of collecting information on the processes within the boundaries of the system under study[3]. The information to be collected typically covers many topics, such as amounts of products, waste, and elementary flows that enter and exit the system, but also other aspects, for example, data quality and uncertainty distribution. LCA studies require a considerable volume of data [4] and their accuracy is crucial to assure the quality of the LCI and LCA study overall [5]. Therefore, data collection and life cycle modelling are widely recognized as challenging steps, for instance regarding multi-output processes and local technology representativeness [6].

LCA practitioners rarely collect all life cycle data themselves, but rather need to combine their own collected data with background information from existing LCA databases [7]. The use of LCA databases enables to (i) reduce time, efforts, and resources for data collection and (ii) reflect supply chains as it would be very difficult for practitioners to depict the whole life cycle of all products in the system [4], [8]. For the latter purpose, the selected database needs to fit to the foreground LCA system. This concerns the geographic and technical representativeness of data contained in databases, among other things. Also, the nomenclature of flows used in the database is crucial. This is why several initiatives to develop national LCA databases have emerged in the past years, in various parts of the world [9]–[12]. Countries have been more and more motivated in performing LCA studies by relying on national databases that reflect their economy, energy mix, and disposal technologies. To develop an LCA database, the creation of LCA datasets is an essential step, which may be hard in the absence of clear guidance and previous related experience. In 2011 the 'Global Guidance Principles For Life Cycle Assessment Databases' – also known as 'The Shonan Guidance Principles'- were published to support LCA database creation worldwide, especially in emerging and developing regions [13]. Since then, several capacity-building projects and workshops have been promoted to enable countries without geographic coverage in common LCA databases to develop country-specific datasets [14], [15]. In addition, the Global LCA Data Access (GLAD) network has been developed starting from 2015 to ease data accessibility and interoperability across different data providers [16].

This article aims to give insights into the main needs, requirements, and challenges for the creation of an LCA database. Specifically, the following questions will be addressed:

- What is needed to create an LCA dataset?
- How to create an LCA database?
- What are the specific requirements and issues for the creation of a national LCA database?
- Are there any tools to support the creation of LCA datasets and databases?

The structure and general approach of the article are presented in the following section.

2. APPROACH

First, the article defines the main characteristics of LCA datasets and discusses data collection approaches to enable the development of the dataset. Secondly, needs, requirements, and issues related to the creation of LCA databases from developed datasets are addressed. Specifically, the case of national LCA databases will be tackled with the aim of discussing lessons learned from past experiences, for instance, decisions to be taken in the development process. Finally, the existence of tools that could ease the LCA dataset and database creation process is investigated, thus resulting in the selection of two tools – the LCA Collaboration Server and the LCA Data-Machine - and presentation of related use cases.

The following chapters on the LCA dataset and database creation and connected tools are developed by combining different sources, ranging from literature research to the professional experience of the authors. Indeed, the authors had the chance to contribute to a diverse set of initiatives in the field of LCA databases, such as the Shonan Guidance Principles [13], the GLAD network [16], and capacity-building workshops in emerging and developing countries [14].

3. LCA DATASET CREATION

An LCA dataset can be defined as the smallest modeling unit in a life cycle model. It is then also called 'unit process'. A life cycle model typically contains many of these unit processes; each of these processes is made of (i) inputs (which are resources and products) and (ii) outputs (emissions, waste, and products), see Figure 1.

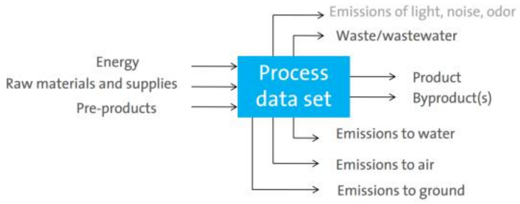

Figure 1: Example of process dataset and its components (resources, emissions, waste, and products)

Of course, every single dataset does not necessarily contain all the different components; for example, wastewater is certainly not an output of every process. The collection of information about each unit process is commonly called 'data acquisition' or 'data collection'. Specifically, it is possible to distinguish between primary and secondary data acquisition.

3.1. Primary data acquisition

Primary data are raw process- and site-specific data collected by the LCA practitioner to create the so-called 'foreground model'. A 'foreground model' is typically understood as made of those 'processes which are under the control of the decision-maker for which an LCA is carried out'[17]. Raw data consist of site- and plant-measurements, estimates, statistics, and bookkeeping and enable the practitioner to compile a list of inputs and outputs for each unit process – or dataset- of the foreground system [13]. In addition, the primary data acquisition procedure foresees the reporting of metadata in view of a dataset creation. Metadata are descriptors providing information on the dataset characteristics, such as the creator name, process name, geographic, temporal, and technology coverage, and format [18]. Therefore, an LCA dataset combines inputs, outputs, and metadata, as shown in Table 1 for the case of primary data acquisition to describe the mango production process in Indonesia[1].

Primary data collection is often still done in a rather 'traditional' way, i.e. by exchanging Excel sheets (see Figure 2) between the LCA practitioner and the different data providers, such as manufacturers and distributors. Furthermore, it often occurs that the data providers identified by the practitioner need to contact other colleagues or actors in the supply chain of the product in order to deliver the requested information. Data collection is certainly to be understood as an iterative process between the LCA practitioner and the data providers.

Table 1: Example of primary data and metadata acquisition for the mango production process in Indonesia, principal data

Mango production - Argentina		
Metadata		
Author	Name Surname	
Year	2018	
Region	Argentina	
Data collection	Primary	
Input		
Agricultural land use	0.5	m2/a
N-fertilizer	0.05	kg
Manure	0.05	kg
Pesticide xy	0.01	l
Irrigation water	2	kg
Biogenic CO2, from air	1	kg
Output		
Mango, fresh	1	kg
Fertilizer, in ground	0.025	kg

Communication and trust play a major role in a successful and accurate data collection. Language clarity and precision when preparing a data collection sheet are crucial aspects to be considered. Understanding of the technical process by the LCA modeler and of the LCA data needs by the technical data provider is key to ask for and receive correct and useful information. It is recommended that the LCA professional works on building a relationship of trust with the data providers, for instance by (i) presenting the purpose of the LCA study and why data are needed from the technical providers´ side, (ii) explaining how the data collection sheet needs to be understood and completed, and (iii) discussing any sensitivity issue and how the results of the study will be used.

Manufacturers and other data providers may not want to deliver data that they perceive as sensitive. Companies are often concerned that disclosing details of their activities could result in negative consequences, such as bad reputation among customers, violation of industrial secrets, and advantages for competing companies. For this reason, data sensitivity is often an issue to be tackled when dealing with dataset creation from primary data.

[1] Data in this table is only used as an example and does not aim to reflect a real process

Protection of sensitive data can be addressed by two different strategies:

- Remove data specificity, by pulling out information that might make the company or plant identifiable. For instance, instead of mentioning the specific location of the plant or chemical used in a process, only the name of the country or if an inorganic or organic chemical is used could be reported.
- Data average and aggregation, by combining multiple datasets to deliver an average dataset representative for a certain product and/or by producing a single dataset where all inputs and outputs of the diverse processes are summed up, of course properly scaled. The result is then an aggregated dataset (in contrast to a unit process dataset).

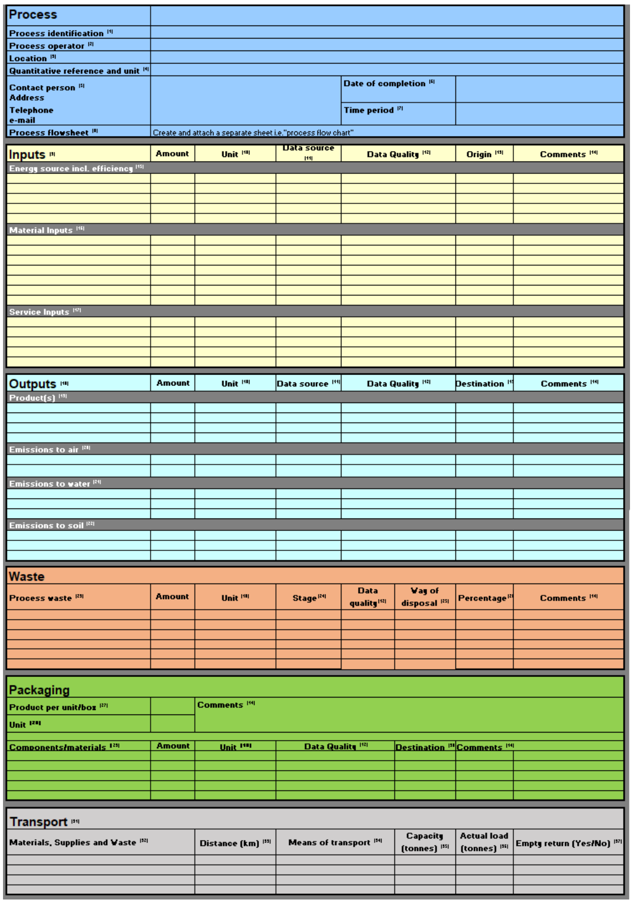

Figure 2: Example of a data collection sheet in Excel

3.2. Secondary data acquisition

Datasets in the foreground system usually do not exceed 5% of all processes in the system under study [19]. The remaining processes constitute the so-called 'background system', 'on which no or, at best, indirect influence may be exercised by the decision-maker for which an LCA is carried out'. The background system is made of secondary data, which are generic information derived by a variety of sources: public and official national statistics, company websites, previous research published in journal papers, and existing LCI databases [20], [21]. In the absence of primary data, also secondary data could be used to complete the foreground model. However, although secondary data and especially background data from LCI databases are essential for any LCA study [7], a number of challenges are often faced by the LCA practitioner when dealing with them:

- Secondary datasets may not be fully representative for the purpose of the study because they are incomplete, for instance, because a product or emission is missing.
- Secondary datasets may not be not fully representative for the purpose of the study because they do not meet some criteria, e.g. geographic, temporal, technological representativeness. For instance, when the practitioner needs data for a production process in Germany, but they are available only for Switzerland.
- Further, secondary data may not fully fit for the modeling choices set for the foreground model and overall LCA study, concerning the treatment of biogenic carbon, of infrastructure, of water flows, allocation rules applied in the datasets, and others
- Reliability of the source needs to be verified to understand whether data reported are accurate and can be trusted. For example, a reliable source should report all assumptions and boundaries of the study from where a certain dataset resulted. It is also recommended to apply the 'triangulation of data' by consulting multiple sources in order to compare and contrast information [22].

Fitness of data for the purpose of an LCA study is a major concern of any practitioner. With this in mind, the GLAD network developed a set of metadata descriptors that could support LCA users in the selection of best fitting data from available background datasets [23]. The set of metadata descriptors includes, among others, process name and type, sample approach, time, geography, technology, verification, and administrative information.

3.3. Creation of datasets towards a database development

To develop a database, many datasets need to be created by using and combining both primary and secondary data. In this context, consistency plays an important role when dealing with modelling decisions, for example regarding End-of-Life and multi-output processes. Coherence in flow use across different datasets is also to be considered, for instance when assigning emissions to processes (e.g. by distinguishing between biogenic and fossil carbon emissions or only using a generic carbon emission) or modelling waste (e.g. by using product or waste flows).

Time- and resource-efficiency are crucial aspects when creating groups of datasets. For this purpose, users may rely on datasets that have already been developed, thus making few adjustments instead of creating the same process twice or starting from scratch. Large efforts are also required for the update of datasets, even more, when the update is intended to be systematic [7]. Improvement of a group of datasets may be prioritized following different logics:

- Data update based on technological development [13].
- Data updates based on market, industry, and public policy progress [13].
- Data update based on the relative process relevance analyzed in a group of datasets for different life cycle impact categories [7].

4. LCA DATABASE CREATION

An LCI database is defined as ‘a system intended to organize, store, and retrieve large amounts of digital LCI datasets easily’[13]. A database should provide (i) comprehensive input/output flows for datasets, (ii) a consistent modelling approach and flow nomenclature, and (iii) complete and consistent dataset documentation. Specifically, consistent modelling among datasets allows the compilation of datasets in a life cycle model and even, in the case of consistency among databases, a combination of different databases. Three main actors can be identified in a database development process [13]:

- The data/dataset provider performs data collection and commits data to the database manager;
- The database manager is responsible for the database development (unit process creation) and management, including storing of datasets, review, validation, and networking of databases.
- Users, finally, use the datasets and provide feedback to the database manager.

For many years, discussions have been held about the “right” format of databases and LCA datasets [24], [25]. For the actual data storage, the format does not appear crucial. However, a database should be able to “deliver” data in the different broadly used formats. This is why discussions have been mainly about the “right” exchange format [25]. At present, probably three LCA data exchange formats are most broadly used: ILCD, developed for the European Commission [26]; EcoSpold02, developed for the ecoinvent centre [27], and JSON-LD, developed for openLCA and used by LCACommons [28]. Since EcoSpold02 is by now not supported by any LCA software, some databases provide datasets in the proprietary CSV format for the SimaPro software, in addition.

Version control is also another important aspect to guarantee, to database users, stability in results. Version control reflects that database managers will need to periodically release updates, fix bugs, and extend databases. These operations need to be documented and be quality assured.

Reliability and consistency of information provided in a database are crucial for the quality of LCA studies and to ensure and maintain users´ trust. Review of datasets can be a powerful contribution to quality assurance and therefore to increase reliability and trust in a database. Key aspects are (i) what is going to be reviewed - scope, (i) reviewer selection, and (iii) review workflow and organization. Performing a review on aggregated processes might not be fully effective as it could be difficult for the reviewer to track and check the modelling decisions and outcomes. Hence, it is recommended to undertake a review of unit process datasets in order to achieve the maximum transparency and quality assurance.

ISO 14071 TS provides a standard for critical review that 'should cover all aspects of an LCA" [29]. However, the assessment of single datasets and LCI models is an optional step in the critical review according to ISO 14071 TS [30]. The Environmental Footprint (EF) initiative by the European Commission [31] provides guidance for performing a dataset review by defining requirements and procedures for validation and verification of EF studies [32]. Ciroth et al. [33] propose 14 review criteria for LCI datasets organized in 5 different clusters: goal, model, value, relevance, and procedure. Unfortunately, a dedicated standard and overall accepted approach for dataset review do not exist. However, several tools to support the review procedure exist, such as EcoEditor [34], the ILCD validation tool [35], and the LCA Collaboration Server (see Section 5.1).

4.1. Seven points for LCI database development

Seven main points for the development of an LCI database can be highlighted.

1. *Database scope and dataset development roadmap*: what is planned to be addressed by the database (time, geography, products) and what are the steps to create the database (which datasets should be created first, how to exchange technical and intellectual information between unit process modelers); it is useful to develop, for one, initial, common datasets for the database that are used in other datasets developed later, the entire database development process then resembles a bootstrapping procedure; on the other side, it is also good to develop "precious" datasets in the database that are interesting to be used thereby increase interest in the database.

2. *Data collection strategies* see Section 3**Error! Reference source not found.**.

3. *LCA methodology and interoperability*: methodology definition is key for a consistent database (for example regarding waste modeling, allocation, flow nomenclature, biogenic carbon modeling, supported impact assessment methods, exchange format) and to allow users to combine datasets from different database providers worldwide. For this latter purpose two approaches can be identified:

   - 'Enforced consistency': harmonization of the methodology in all databases. This raises questions like who decides about the correct methodology, what are different use cases that call for different methods, and how to ensure real compliance against fake consistency (i.e. pretending to be consistent).

   - Implement interoperability: make databases with different methodologies usable. The GLAD network [16], for example, aims at implementing metadata descriptors to evaluate the best available datasets from different providers, see Section 3.2. The idea is to let users specify what they are looking for and provide suggestions based on the metadata descriptors that could be implemented in a database search [23].

4. *Quality assurance for the database*, see section 4.

5. *IT infrastructure* for the database to support the development, maintenance, sharing, and update of the database. For example, see section 5.1 on the LCA Collaboration Server.

6. *Business model and maintenance* of the database, including coverage of running costs, costs for maintenance, and upgrades to keep the database relevant. Databases can be developed and maintained (i) with full public support, (ii) as full commercial databases, and (iii) with public support and income from license sales.

7. *Making the database used*, as it will be rarely used stand-alone, but rather in LCA software and tools. Therefore, it is recommended to engage in

discussions and collaboration with tool developers and users when creating the database.

4.2. Developing National LCA databases

A national LCA database is a database with the aim to provide a reference for LCA studies that address products and services occurring in the specific geography covered by the database, typically a country. LCA databases were first developed for many European countries, the United States, Japan, and Australia, and also with the ELCD database, for the European Union. In other regions, processes are often very different from those contained in the first developed databases; this pushed the need for additional national databases, which are often provided for free and maintained by public institutions. Examples of national databases include the cases of Chile, Brazil, Malaysia, Thailand, India, and Sri Lanka. When creating national LCI databases, stakeholder discussions among LCA software providers, consultants and researchers, and the national industry are crucial. Indeed, results from national databases and studies that use those databases are interesting for the national industry. Furthermore, the market entry of novel national databases can have a high impact, for example by affecting the use of other existing databases.

Sometimes, national databases have the goal to provide "reference" datasets that can be used as a reference for environmental impacts of products, or of process datasets, in a given region or country. An example is the ELCD database, which even was called "European Reference Life Cycle Inventory Database" [36]. In rare cases, national databases are foreseen to provide legally binding information. One example is Switzerland, where biofuel subsidies depend on an overall sound carbon balance for the supply chain of the biofuel which is to be calculated using the Swiss ecoinvent database [37].

National LCA database creation is probably more challenging than the creation of other LCA databases; the development implies decisions about, inter alia, the seven points listed in the preceding section, but maybe more complicated for several reasons:

- With the idea to provide reference data which potentially even are meant to provide legally binding information, the need for quality assurance and reliability is higher than for normal LCA databases
- National databases are perceived as a large project with high relevance, which simply causes more stakeholder reactions, from industry and also from consultants and LCA database providers; this can make also rather technical decisions about flow nomenclature, for example, long and tedious
- Sometimes, maintenance of the database is more difficult, since the idea is often to provide the reference database for free, which makes maintenance and management dependent from other sources of budget, typically public ones
- whether the database should contain unit processes, system processes, or even LCIA results only;

There are some examples where national databases did not successfully start (a German database planned within a German network for LCA data, for example), or were discontinued after some time. An example of the latter is again the ELCD database [38].

As a consequence, it seems, for one, wise to develop a national database not in one step but instead from a "precursor" database that also serves to gain experiences. And second, for creating a national LCA database, stakeholder inclusion and interaction, quality assurance and creation of reliable datasets, and development of a sound business model with long term "sustainability", deserve special attention.

5. TOOLS FOR LCA DATABASE CREATION

As outlined in the previous sections, developing LCA datasets and creating an LCA database are tasks that require time and accuracy. As also mentioned above, the development of datasets is often done in a rather traditional way, which can lead to time demands of more than three months for only entering less than 50 datasets into a database. It makes thus sense to seek for ways to make the database creation faster and more efficient. One promising way is modern tool support. Two tools are presented in this chapter, the LCA Collaboration Server (CS) and the LCA Data-Machine, both developed recently by GreenDelta.

5.1. LCA Collaboration Server (CS)

The LCA CS [39] is a server application that complements the software openLCA [40]. It has been designed with the objective to facilitate the exchange of LCA datasets between users and was initially commissioned by the US Department of Agriculture, in the context of the LCA Digital Commons [41]. It is free software and can be obtained from the openLCA download page (openLCA.org). The CS empowers users to conduct distributed and simultaneous collaborative LCA modelling by enabling (i) synchronization of databases, (ii) tracking of changes, and (iii) comparison of databases. The tool ensures the connection between a local database in the openLCA software and an online repository which is equivalent to a database in openLCA and contains the same elements; a repository mirrors the local database of the users connected to it. Users can 'commit' (i.e. push) data from a local openLCA database to an LCA CS repository as well as 'fetch' (i.e. download) data from a repository to a local database. For one CS, it is possible to define several users and teams linked with one or several repositories; different rights and roles can be assigned to groups and users.

Different cases can be identified for the use of the CS as a supporting tool for LCA dataset and database creation, including:

- Iterative co-development of datasets and LCA models, data collection between manufacturers, and LCA practitioners (see Figure 3).

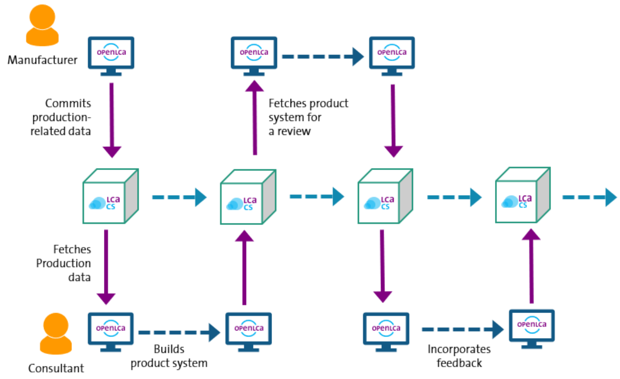

Figure 3: CS use case: iterative co-development of datasets and LCA models

- Building and managing verified public LCA repositories, for instance, a national LCA database (Figure 4).

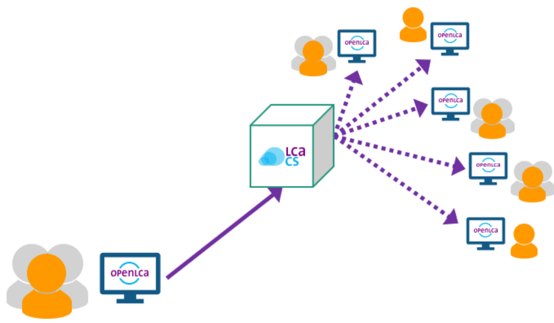

Figure 4: CS use case: building and managing verified public LCA repositories

- Publication: Straightforward sharing of LCA models, flows, processes, and entire databases, for instance for the creation and publication of a national database from different contributors (Figure 5).

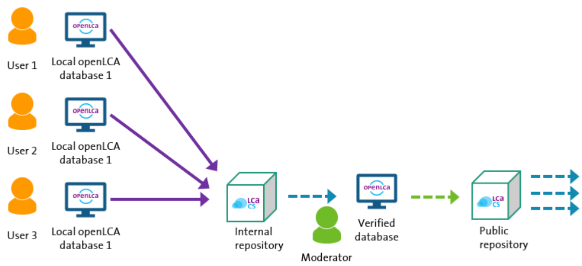

Figure 5: CS use case: database publication

5.2. LCA Data machine

Under development for a German research project, the LCA data-machine (DaMa) is a tool to create datasets for LCA that is (i) demand-driven, (ii) automatic and fast, and (iii) with a controlled fitness for purpose. The DaMa aims at gathering and combining various data from different sources, from the LCA domain and outside of the LCA domain, in different formats in an accessible and understandable way for everyone.

The DaMa is based on the following principles:

- Creation of a data pool and structure from various sources. A data pool is a compilation of data from different LCI databases, emission repositories, Environmental Product Declarations, and other sources. The structure is represented by 14 process archetypes; every product belongs to an archetype. For instance, in the case of agriculture, “Production of materials” is an archetype that includes subprocesses such as “Combustion”, “Material conversion, industrial”, “Agricultural production, plants”, “Transport, pipelines”, and “Goods transportation”.

- Definition of rules based on mathematical principles in order to identify better-fitting data for users´ needs and applications. Rules for data are based on three dimensions, (1) reliability, (2) preciousness (meaning how data were obtained, such as expert judgement, estimate, and measurement), and (3) fitness for purpose.

- Design of a user interface to generate datasets. This forms the output tool where the user can select the archetype and refine it with a more specific product description (Figure 6).

- Generation of datasets that fit users’ requirements and can be exported in JSON-LD format (see Figure 7). The final result is a list of inputs and outputs where the amounts are inter-related, with an assessment of the fitness, and with metadata.

The DaMA should not be seen as in competition to the existing LCA data sources, but rather as an addition to satisfy a broader need, to create datasets faster and for more widespread uses. Especially, this tool can have a high potential for fast and systematic creation of datasets to be combined in a national LCA database. With this perspective, the DaMa is currently under testing for the creation of a national LCA database for India.

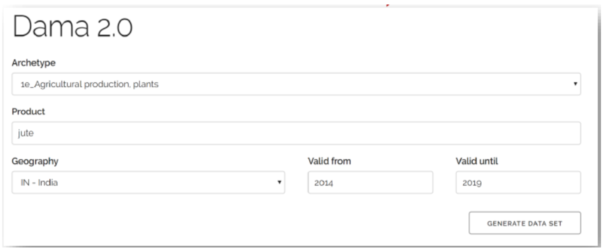

Figure 6: User interface of the LCA Data-Machine

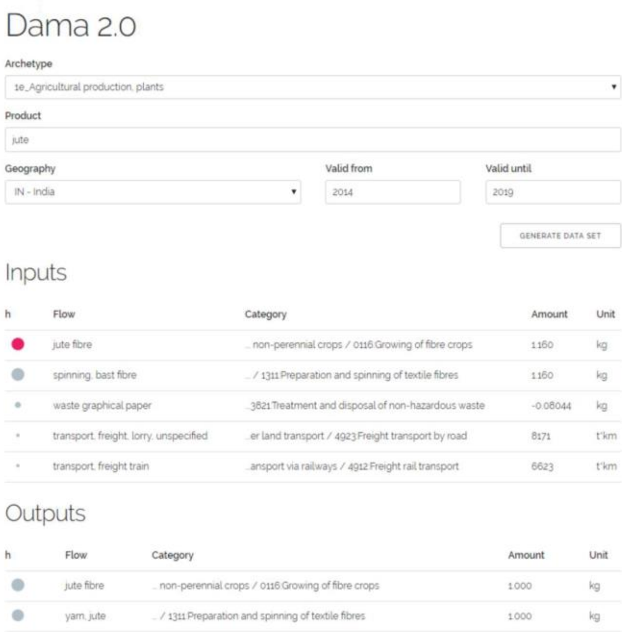

Figure 7: Example of a dataset generated by the LCA Data-machine

6. CONCLUSIONS AND FURTHER OUTLOOK

The article outlines the main aspects related to the LCA dataset and database creation. Development of datasets requires time, effort, and resources in order to collect and combine primary and secondary data. Having a database to model the background system and the supply chain of a product or service eases the process of life cycle modelling and allows the user to focus on primary data collection for the foreground system. The availability of datasets fitting the purpose of the study is crucial for the accuracy of the results and the quality of the LCA study. LCA can become a widely spread methodology for decision-making only if results are accurate and reliable, hence if the underlying data and life cycle modelling steps can be trusted.

Geographic coverage is probably one of the first parameters to be considered to evaluate data quality. A dataset should reflect as much as possible a process under study and its upstream life cycle chain. In this perspective, generic European or American databases might not be suitable to reflect processes occurring in other regions and products typical to specific countries: better fitting regional, and potentially even national LCA databases are needed. A number of national databases have already been developed and more initiatives will be undertaken in this direction. It is important that countries willing to create a national database are supported by actors and institutions with a long tradition in the field, for example with capacity-building workshops and professional exchanges. Access to free, or partially free, LCA tools and IT infrastructure is key for database development in developing and emerging countries. With a national LCA database in place, those countries can have an additional tool to orient their economy to sustainable production and management pathways, be more competitive in international markets if specific sustainability or certification standards are required, and contribute to job creation, for example as LCA researchers and consultants. Indeed, also countries with a long tradition in LCA will benefit from interactions with newcomers, for instance by discussing together unsolved methodological and interoperability issues, and by testing new approaches and tools. Creating national databases fosters national industry and science. This is a chance for local LCA communities to bring in innovation and benefit from existing experiences at the same time.

REFERENCES

[1] International Organization for Standardization, “ISO 14040:2006._Environmental Management - Life Cycle Assessment - Principles and Framework_. ISO; 2006,” 2004. doi: 10.1016/j.ecolind.2011.01.007.

[2] International Organization for Standardization, “ISO 14044:2006._Environmental management — Life cycle assessment — Requirements and guidelines_ ISO; 2006,” 2016.

[3] European Commission -- Joint Research Centre -- Institute for Environment and Sustainability, *International Reference Life Cycle Data System (ILCD) Handbook -- General guide for Life Cycle Assessment -- Detailed guidance*. 2010.

[4] G. Bourgault, P. Lesage, and R. Samson, “Systematic disaggregation: a hybrid LCI computation algorithm enhancing interpretation phase in LCA,” *Int. J. Life Cycle Assess.*, vol. 17, no. 6, pp. 774–786, Jul. 2012, doi: 10.1007/s11367-012-0418-7.

[5] N. Meron, V. Blass, and G. Thoma, “Selection of the most appropriate life cycle inventory dataset: new selection proxy methodology and case study application,” *Int. J. Life Cycle Assess.*, vol. 25, no. 4, pp. 771–783, 2020, doi: 10.1007/s11367-019-01721-8.

[6] J. Reap, F. Roman, S. Duncan, and B. Bras, “A survey of unresolved problems in life cycle assessment. Part 1: Goal and scope and inventory analysis,” *Int. J. Life Cycle Assess.*, vol. 13, no. 4, pp. 290–300, 2008, doi: 10.1007/s11367-008-0008-x.

[7] J. Reinhard, G. Wernet, R. Zah, R. Heijungs, and L. M. Hilty, “Contribution-based prioritization of LCI database improvements: the most important unit processes in ecoinvent,” *Int. J. Life Cycle Assess.*, vol. 24, no. 10, pp. 1778–1792, 2019, doi: 10.1007/s11367-019-01602-0.

[8] J. Reinhard, C. Mutel, and G. Wernet, “Contribution-based prioritization of LCI database improvements: method design, demonstration, and evaluation,” *Env. Model Softw*, vol. 86, 2016, doi: 10.1016/j.envsoft.2016.09.018.

[9] K. Chomkhamsri, T. Mungcharoen, and C. Yuvaniyama, “10-year experience with the Thai national LCI database: case study of ‘refinery products,’” *Int. J. Life Cycle Assess.*, vol. 22, no. 11, pp. 1760–1770, 2017, doi: 10.1007/s11367-016-1160-3.

[10] X. GONG, Z. NIE, Z. Wang, and T. ZUO, “Research and development of Chinese LCA database and LCA software,” *Rare Met.*, vol. 25, no. 6, Supplement 2, pp. 101–104, 2006, doi: https://doi.org/10.1016/S1001-0521(08)60061-3.

[11] I. Vázquez-Rowe, R. Kahhat, and I. Sánchez, “Perú LCA: launching the Peruvian national life cycle database,” *Int. J. Life Cycle Assess.*, vol. 24, no. 11, pp. 2089–2090, 2019, doi: 10.1007/s11367-019-01668-w.

[12] P. Lesage and R. Samson, “The Quebec Life Cycle Inventory Database Project,” *Int. J. Life Cycle Assess.*, vol. 21, no. 9, pp. 1282–1289, 2016, doi: 10.1007/s11367-013-0593-1.

[13] UNEP/SETAC, “Global Guidance Principles for

Life Cycle Assessment Databases - 'Shonan Guidance Principles,'" UNEP/SETAC, 2011.
[14] Sri Lanka, "Development of National LCA Database Roadmaps – NCPC," 2019. .
[15] "The REAL project | UN Environment," 2016. .
[16] "GLAD." .
[17] R. Frischknecht, *Life cycle inventory analysis for decision-making: Scope-dependent inventory system models and context-specific joint product allocation*, vol. 3, no. 2. 1998.
[18] M. Lohrey and M. Librarian, "LCA Metadata: Challenges and Opportunities," doi: 10.2788/9588.
[19] B. Steubing, G. Wernet, J. Reinhard, C. Bauer, and E. Moreno-Ruiz, "The ecoinvent database version 3 (part II): analyzing LCA results and comparison to version 2," *Int J Life Cycle Assess*, vol. 21, 2016, doi: 10.1007/s11367-016-1109-6.
[20] GreenDelta, "openLCA Nexus: The source for LCA data sets," *Databases*, 2019. .
[21] R. Frischknecht *et al.*, "The ecoinvent Database Introduction 3 The ecoinvent Database: Overview and Methodological Framework," doi: 10.1065/lca2004.10.181.1.
[22] C. Benoît and B. Mazijn, *UNEP/SETAC Life Cycle Initiative - Guidelines for Social Life Cycle Assessment of Products*, vol. 15, no. 2. UNEP, 2009.
[23] A. Ciroth, P. Arbuckle, E. Cherubini, C. M. L. Ugaya, and A. Edelen, "Task 3: Core meta-data descriptors and guidance on populating descriptors WG3 of the Global Life Cycle Data Access Network (GLAD)," 2017.
[24] J. Guinée, "Handbook on life cycle assessment - Operational guide to the ISO standards," *International Journal of Life Cycle Assessment*, vol. 6, no. 5. p. 255, 2001, doi: 10.1007/BF02978784.
[25] I. Meinshausen, P. Müller-Beilschmidt, and T. Viere, "The EcoSpold 2 format—why a new format?," *Int. J. Life Cycle Assess.*, vol. 21, no. 9, pp. 1231–1235, 2016, doi: 10.1007/s11367-014-0789-z.
[26] European Commission, "European Platform on Life Cycle Assessment - Developer." .
[27] Ecoinvent, "ecoSpold2 – ecoinvent." .
[28] OpenLCA, "openLCA Schema." .
[29] International Organization for Standardization, "Environmental management — Life cycle assessment — Critical review processes and reviewer competencies: Additional requirements and guidelines to ISO 14044:2006," 2014.
[30] M. Z. Hauschild, R. K. Rosenbaum, and S. I. Olsen, "Life Cycle Assessment: Theory and Practice," *Life Cycle Assess. Theory Pract.*, pp. 1–1216, 2017, doi: 10.1007/978-3-319-56475-3.
[31] European Commission -- Joint Research Centre -- Institute for Environment and Sustainability, "Product Environmental Footprint (PEF) Guide; CONSOLIDATED VERSION," 2012.
[32] European Commission, "Product Environmental Footprint Category Rules Guidance," *PEFCR Guid. Doc.*, p. 238, 2018.
[33] A. Ciroth, C. Foster, J. Hildenbrand, and A. Zamagni, "Life cycle inventory dataset review criteria—a new proposal," *Int. J. Life Cycle Assess.*, vol. 25, no. 3, pp. 483–494, 2020, doi: 10.1007/s11367-019-01712-9.
[34] "ecoEditor – ecoinvent." .
[35] "Soda4LCA. ILCDValidationTool." .
[36] D. Garraín, S. Fazio, C. De La Rúa, M. Recchioni, Y. Lechón, and F. Mathieux, "Background qualitative analysis of the European Reference Life Cycle Database (ELCD) energy datasets-part I: fuel datasets," 2011, doi: 10.1186/s40064-015-0915-9.
[37] MinöStG, "SR 641.61 Mineralölsteuergesetz vom 21. Juni 1996 (MinöStG)," 1996. .
[38] "ELCD." .
[39] GreenDelta, "LCA Collaboration server | openLCA.org." .
[40] Greendelta, "openLCA.org | openLCA is a free, professional Life Cycle Assessment (LCA) and footprint software with a broad range of features and many available databases, created by GreenDelta since 2006," 2017. .
[41] Federal LCA Commons, "Enter the Commons... | Life Cycle Assessment Commons." .