LCA Methodology

LCI Data Modelling and a Database Design

Raul Carlson, Anne-Marie Tillman, Bengt Steen, Göran Löfgren

TEP - Technical Environmental Planning, Chalmers University of Technology, S-41296 Göteborg, Sweden
Nordic Port AB, S-41265 Göteborg, Sweden

Corresponding author: Raul Carlson

Abstract

A large scale operative data format for transparent storage, administration and retrieval of environmental Life Cycle Inventory (LCI) data has been implemented by applying data modelling and database design.

Key concepts in the design are "activity" and "flow": An activity is a technical system, such as a process or a transport, or an aggregate of different processes or transports. A flow is any matter entering or leaving an activity, such as natural resources, energy, raw material, emission, waste or products.

Any numerical data set on an activity can be thoroughly described by supplying meta data. Meta data fields are prepared for a wide set of commonly known LCA-data aspects, such as descriptions of data acquisition methods, system boundary conditions and relevant dates.

Keywords: Computable data; database design, activity; database design, flow; LCA data aspects; LCI data modelling; LCI; Life Cycle Inventory (LCI); meta data fields; SPINE; Sustainable Product Information Network for the Environment (SPINE); system analysis; transparent storage

1 Introduction

This paper presents a flexible Life Cycle Inventory (LCI) data model, which has been mapped and implemented into a relational database table structure named SPINE (Sustainable Product Information Network for the Environment) [1]. The data model supports full LCI data transparency, both in regard to reference data in the same database and in regard to reference data sources external to the database. The model is methodologically flexible since it supports any LCI methodology, and it is technically flexible since it is not restricted to being implemented as a relational database, but can be implemented using any existing database technology. The data model is also useful as a design aid when designing calculating LCI and Life Cycle Assessment (LCA) software tools, or when modelling and designing LCI data communication formats. The choice of implementing SPINE as a relational database is in itself a flexible choice, since a relational database can be used in many technically and practically different ways. Therefore SPINE can be used as a storage medium for LCI data and LCI reports or it can be used as a data storage format for LCI or LCA software or it can be used as both, alternately or simultaneously.

SPINE is the merged outcome of two separate projects. The aim of these projects was to develop a database that could be used both as a data storage medium for LCI data and as a data storage format for LCA software [2,3]. The first of these two projects, a master thesis, focused mainly on finding a data model which satisfied the general and specific needs formulated by the LCI methodology and which, at the same time, was correct from the viewpoint of computing science and database technology. The other project, Nordic Project on Environmentally Sound Product Development (NEP) [1] (as described below), focused on reaching an industrial consensus in regard to a datamodel and the required contents of an LCI database, indiscriminate to economical or industrial sector. Concerning data modelling, the latter project focused on feasibility and applicability.

During both projects, LCA practitioners and system developers worked closely together in a continuous dialogue, exchanging their respective expertise. This cross fertilisation of competence was an important contributor to the outcome of the two projects and to their final merging into SPINE.

The project performed as a master thesis, was a co-operation between the departments of Technical Environmental Planning and Computing Science at Chalmers University of Technology, Göteborg (1994) [4]. The industrial consensus project was a sub-project within the NEP project, aiming at finding a common LCA database format for Scandinavia (1994-1995) [1].

Twenty-two large companies and institutes in Scandinavia participated in NEP.
2 Related Works

LCI databases are important in practical work with LCA. This because it is difficult to find suitable data, and when data is found it often originates from different sources, has different formats, and must be interpreted and reformatted into a form that suits LCA. This makes data retrieval both an expensive and a time-consuming part of LCA. Databases are expected to solve many of these problems and, consequently, a great deal of work has already been performed in this area, as may be exemplified by the great interest that has been shown towards the SPOLD data format work, as well as towards a great many earlier works:

- Major database collections of LCI data:
  1994: Swiss database for LCI on energy systems [5]

- Serious approaches towards recommendations and requirements for LCA databases:

- Projects aimed at finding a co-ordinated common view of LCA data:
  1996: An approach to creating a common format for Life Cycle Inventory Data, SPOLD [9]

- Major approaches toward computer-based databases for storing easily retrieved LCA data:

- Databases as files within LCA calculation software:
  Examples: LCA Inventory Tool [3,12] and SimaPro

More approaches to LCA databases may be found. However, it should be stressed that none of these works have had a general data modelling approach towards the task. Therefore, these solutions are typically either methodologically imprecise (BOUSTEDT'S General principles... or EKVALL'S Criteria...) or are they too rigid (typical for implemented systems like LCA Inventory Tool and SimaPro or early standardisation approaches like SPOLD's data format).

3 The Approach

The intention was to develop a database that transparently should store LCI information on any technical system, of any size or type, such as single machines or entire production networks following a product from cradle to grave. It should be possible to store information on both single technical systems, as well as on models of technical systems assembled from information on other technical systems in the same database. The latter should enable the storage of fully transparent reports by literally including the contents of the references and their interpretation within the report of any specific study. In addition to reusing information on simple original data sets, this should also enable the reuse of entire models of technical systems. This should be practically useful for information on supply systems, for example, electrical power, waste treatment and wastewater treatment.

It was intended that the database should support any LCI approach or methodology, e.g. LCIs with different approaches to system boundaries or allocation methods, and it should be general in terms of database technology. In addition to this, it should be possible to use the same database as a data storage format for different LCA software, as well as for general databases for LCI data.

This approach gave the following criteria for the database:

1. In order to conceptually define the term "technical system", the database must be based on a general model.
2. In order to enable full data transparency, the database must allow for descriptive information about each technical system and about any quantitative data in the database. It should be possible to describe every data entity for which ambiguity may be raised.
3. To make the result generally available and independent of any specific database technology, a conceptual data model of the information needed to describe LCI data was needed. It should be possible to implement this model into any database system.
4. The intended generality regarding the use of the database implies requirements on the choice of implementation: the database should be based on generally accepted, computer-based database technology.

SPINE complies to these criteria by a series of solutions: In the following section (Modelling Transparent LCI Data), a conceptual systems model\(^1\) will be described which enables SPINE to store any size and any type of a technical system. In the subsequent section (LCI Data) a description of the data needed to describe these systems is provided.

Having made a conceptual model of LCI data implies that the database solution will be independent of any specific database system. The requirement for the database to be useful both as a storage format for LCA software and for large LCI databases suggested the use of a relational database system. This because they are commercially available and because the competence on this technology is spread widely.

\(^1\) The conceptual model provides an interpretation to the concepts and their relations.
4 Modelling Transparent LCI Data

When data is used by and communicated within a large group of users, for their work and for their decisions, there is a need for a common understanding of the data. Not only must the data format be understood collectively, but all relevant and implicit understanding of the data must also be made explicitly, and be supplied together with the data. Explanations cannot be adhered orally, and cannot be implicitly understood from the familiarity with the contextual situation in which the data originally evolved.

Data modelling is a means for formalising data shared within a group of information users. The purpose of the data modelling process is to retrieve all relevant types of information on a certain topic, and to seek to find an effective and efficient image (model) of that information. The result is aimed to allow any user within the group all the information which can be retrieved from within the entire group and which this user needs to fulfil his or her responsibility within the group. That is: anyone who needs information can reach it, while it can be secured from anyone who should not need it.

When modelling SPINE, the modelling topic was LCI, as a part of LCA, and the group of users was practitioners, reviewers and commissioners of LCA.

4.1 The concept of activity

In LCA, the detailed operations of a process are of limited interest. The items of main interest in LCA are the inputs to and outputs from the processes. In this respect, transport processes, transforming processes and other processes have enough similar characteristics to be modelled as one common concept, which we call activity.

For most processes, the inputs and the outputs are dependent, linearly or not. The data model as presently implemented, supports only linear relations between inputs and outputs, although the conceptual model supports any dependencies. Transports, however, differ from most other activities, in that the inputs and outputs also depend on another parameter, the transport distance. Thus, it was found useful to differ between activities with outputs depending on inputs only, and those where an additional parameter is needed to calculate the inputs and outputs. Figure 1 graphically exemplifies the modelling transformation from different types of processes into activities.

Most processes can be disaggregated into sub-processes. For example, a process for making steel sheet (→Fig. 2) can be disaggregated into the process of making steel bars from iron ore, rolling the bars into thick sheets, and then rolling the thick sheets into thin sheets. The steel making process can then be disaggregated further into a primary steel making process and a casting process. A disaggregation like this can generally be applied to any complex process.

In LCA, a process disaggregation is used in the goal definition stage, or during the inventory stage: The initial technical system, supplying a product or a functional unit, is divided into a raw material extraction activity, a manufacturing activity, a use activity and a waste management activity. Through further stepwise refinement, the raw industrial activities may be seen, with a successively increased level of detail.

The possibility of aggregating and disaggregating activities was modelled by extending the meaning of the activity concept: An activity is either aggregated, which means that it has internal activities, or it is a simple activity.

The aggregated activity is the key to allow for transparent LCI reporting, reuse of plain data on technical systems and reuse of entire models of technical systems:

- By internalising a plain activity within an aggregated activity, the entire set of references and explanations is also internalised, hereby guaranteeing absolute transparency.
- A plain activity can be reused within an unlimited number of aggregated activities.
- Since both plain and aggregated activities are treated identically, an aggregated activity may include any number of both plain activities as well as aggregated activities.

Here "real" means a level of process aggregation for which data is accessible, sometimes referred to as "unit processes" [9].
4.2 The concept of flow

It is not possible to anticipate all future needs regarding types of inputs and outputs. Some typical types are: "raw material", "electric energy", "fossil fuel", "product", "waste" and "emission". Depending on how and why an LCI is made, there may be a need to include different numbers of types. For example, it may be necessary to distinguish between "virgin" and "recycled" raw material, or "main products", "by-products" and "waste". For this reason it would be unnecessarily inflexible to design a conceptual model for some fixed set of types. Instead, we modelled only two general types, named "inflow" and "outflow". The two are conceptually referred to as flow, and they can be of any subtype, in concert with a purpose.

Any activity has an unlimited number of inflows and outflows.

Identification of activities to include in the technical system of the inventory is performed by starting from a specific product or service. From this starting point, the matter and energy paths are followed, upstream to the cradle of the raw material and downstream to waste treatment. The result is a technical system that is not merely a collection of activities and subsystems, but rather a network of activities. Each activity is incorporated into the technical system on the basis of its exchange of matter or energy with the previously incorporated activities.

A network is built by connecting flows of activities within an aggregated activity (→ Fig. 3).

Since resource use, emissions, waste and by-products are also flows, there will be a number of flows not connected to any other flows of other internal activities. These free flows represent resource use, use of raw material, emissions, waste and by-products out of or into the total technical system of the study: The flows of an aggregated activity are the sum of the free flows of its internal activities (→ Fig. 3).

It is not necessarily true that all activities are connected within a technical system. For example, when allocation is avoided by applying system expansion [20], it may lead to technical systems consisting of two or more separated technical subsystems.
The fact that an aggregated activity can have flows, as can any other activity, makes the conceptual model of activity and flow extremely general: any technical system can be transformed into an activity, and can therefore be incorporated into any other technical system. The fact that an aggregated activity can have flows qualifies it to participate in flow connections, which means that it can exchange both matter and services with other internal activities inside an aggregated activity.

The flow concept allows an activity to have any number of products. The data model also supports any allocation method applied to such a production system. This support has been implemented differently for different allocation methods.

The conceptual activity model inherently supports avoidance of allocation, by system expansion or by an increased level of detail as suggested in 1996 in the ISO draft CD 14041.2 [13]. Any extra technical systems can be added to an aggregated activity, and any single activity can be exchanged with an aggregated activity to increase its level of detail.

A general method for both non-linear and linear allocations is also supported by the model: Allocation based on non-linear process behaviour is not supported in the present implementation of the data model, but it might very well be expanded to support also this.

Linear allocation based on, for example, relative economical value is supported fully in two ways: Any substance and any flow may be subscribed an arbitrary number of quantitative properties, for example economical value. Thereby giving possibility to register the data needed for linear allocation decisions. In addition to this, it is also possible to register linear, quantitative relationships between any pair of flows of an activity internal of an aggregated activity.

By having defined the activities as systems, any principle of allocation may be applied to them, either automatically by rules implemented in software, or by manually defining a new allocated activity which references the original data set. Also, it is possible to describe any allocations that may have been applied. This, of course, is especially useful for data without references within the database, since many decisions must be made when isolating any technical system from its (technical and economical) environment. This is discussed further below while describing the meta data abilities of SPINE.

5.1 Computable data

Amounts of flows are computable data, as is transport distances and relative economical value.

There are many possible ways to represent amounts. The choice depends both on how the retrieved data is formatted and on whether or not the normalisation of a technical system includes statistical calculations. The possible formats prepared for in the present implementation of SPINE is: Single numerals (quantity = 35.2), assumed Normal-distributed mean values ($\bar{x}$) supplied with a variance $s$ (quantity $= 35.2$, $s = 4$) or any other type of statistically interpretable value, supplied with a minimum and a maximum value in accordance with some statistics for the values (quantity $= 35.2$, min $= 28.3$, max $= 41.0$).

However, we are fully aware that this approach will need to be further developed in pace with an emerging use of statistical methods in LCA methodology, and an emerging growth of the statistical understanding of retrieved LCI data.

5.2 Meta data

The concept of activity and flow is an image, or a model, which may be communicated by itself in order to synchronise an implicit and collective understanding of computable LCI data within a group of data users and suppliers. However, due to inventory subjectivity, data gaps, system boundary difficulties and other pitfalls of LCI, there is also a need to supply data with explanatory data, meta data.

In general, there are two types of meta data needed for LCI data:

- **The physical object**: A relevant description of the physical object which data is meant to represent.
- **The principles, procedures, conditions and decisions applied and obeyed**: Relevant descriptions of the principles, procedures, conditions and decisions applied and obeyed when retrieving and translating information into the present data representation of the physical object.

These two meta data types will be further described in the following two subsections.

5.2.1 Describing physical objects

Activities may have similar names and produce similar products, but may still be environmentally different due to differences at the system level. These differences may be found in the internal functional subsystems or in the choice of system boundaries (technical, geographical, time, ...). Therefore, such information needs to be supplied with the data.

*We use the term "computable" in a contextual sense, meaning data computable within the LCA methodology.

*"Meta" – from the Greek meaning "transcending or going beyond". Used here to mean "Data about data".*
Examples of relevant information on the internal of systems are shown in Table 1.

Examples of system boundary conditions that needs to be described are exemplified in Table 2.

Table 1: Internal aspects of technical systems which may relevantly affect their environmental behaviour in relation to a similar system producing the same product

<table>
<thead>
<tr>
<th>Internal aspect of system</th>
<th>Example</th>
<th>Example in general terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factual technology</td>
<td>The name of the technology applied to produce a product</td>
<td>It may exist more than one alternative production technology</td>
</tr>
<tr>
<td>Emission reducing devices</td>
<td>Filters or waste water treatments are present</td>
<td>Type and location for the environmental impact has been changed</td>
</tr>
<tr>
<td>Technological improvements</td>
<td>Decreased temperature for chemical reaction to occur</td>
<td>Changes made to a standard procedure or process, thereby reducing use of resources</td>
</tr>
<tr>
<td>Energy saving devices</td>
<td>Heat exchangers are installed</td>
<td>Waste has been turned into goods by technological innovation</td>
</tr>
</tbody>
</table>

Table 2: System boundary aspects that may affect the environmental impact of the technical system

<table>
<thead>
<tr>
<th>System boundary condition</th>
<th>Example</th>
<th>Example in general terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical system boundaries</td>
<td>Administrational supply and internal transports</td>
<td>Included or excluded system components</td>
</tr>
<tr>
<td>Geographical system boundaries</td>
<td>Geographical site of electric energy generation</td>
<td>The region within which a technical system is extended, and the surrounding, not included regions</td>
</tr>
<tr>
<td>Time boundaries</td>
<td>The time for which the data set is relevant</td>
<td>Any time-related aspects of the system that may affect its current state</td>
</tr>
</tbody>
</table>

The implementation of SPINE allows for structured storage, retrieval and administration of this descriptive information, together with the computable data.

Material and substance flows must also be described, in order to unambiguously identify them. This is mainly due to naming ambiguity. For example, "Iron" could mean "Iron ore as a natural resource", "Mined iron ore", "Cast iron", or even "Steel" or "Steel product". Therefore, vaguely defined names, together with vaguely described technical systems, may lead to erroneous LCA results, due to double accounting or an unnoticed lack of subsystems.

The SPINE model provides structured support for implementing and applying any hierarchical nomenclature for material and energy: In order to give name to a flow, the name must first be inserted into the hierarchical substance nomenclature. Thereafter, the name can be chosen for the flow. This arrangement ensures a second thought when naming data, and it provides a practical means for standardising naming within groups of LCA users: Data suppliers can be supported with ready made and simply used nomenclatures within their database, thereby applying the same foundation for naming.

5.2.2 Describing principles, procedures, conditions and decisions

A set of data representing a technical system is the result of applied principles and procedures in compromise with subjective decisions due to factual conditions: The technical system must be selected: a specific plant, for example, or an average formed from a population of similar plants. It must also be decided which parameters to include, how to draw the system boundaries, how to deal with time, etc.

It is not likely that any two representations will be the same if created by different people under different conditions. Therefore, in order to correctly understand and interpret the data, it is necessary to supply the data with a description of the decisions that led to the results, as well as of the people responsible for the result, and of the conditions under which the data was put together. SPINE has been designed to structurally allow for such descriptions on any data set.

Examples of how the principles and conditions can be applied when defining a system boundary is given in Table 3.
Computable data of course is also the result of compromises between principles and facts. Examples of descriptions that may be supplied with each separate computable data entity are shown in Table 4.

Table 3: System boundary aspects and the subjective aspects resulting in these boundary choices

<table>
<thead>
<tr>
<th>System boundary aspect</th>
<th>Example</th>
<th>Subjectivity aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical system boundaries</td>
<td>Applied allocation or cut off rules and assumption regarding their relevance</td>
<td>Applied principle for allocation and compromised application of this principle under actual condition</td>
</tr>
<tr>
<td>Geographical system boundaries</td>
<td>Description of assumptions regarding geographical extensions of technical support systems</td>
<td>Applied principle for cutting support systems at national borders and compromised application of this principle under actual condition</td>
</tr>
<tr>
<td>Time boundaries</td>
<td>Description of assumptions regarding the time for data relevance</td>
<td>Applied future scenario method and compromised application under actual conditions</td>
</tr>
<tr>
<td>Environmental boundaries</td>
<td>Description of applied criteria for selecting the relevant and significant flows</td>
<td>Applied impact assessment model for identifying the relevant parameters and compromised application of this model under actual condition</td>
</tr>
</tbody>
</table>

Table 4: Aspects of computable data that are subject to subjectivity and therefore should be supplied with a description

<table>
<thead>
<tr>
<th>Aspect of computable data</th>
<th>Example</th>
<th>Subjectivity Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>The origin of the computable data entity</td>
<td>How and when they were measured, or otherwise retrieved</td>
<td>Identification of empirical value of the computable data entity</td>
</tr>
<tr>
<td>The statistics of the computable data entity</td>
<td>How they were mathematically and/or statistically treated</td>
<td>Identification of the scientific value of the methods applied to the computable data entity as well as the value of the statistics provided with the entity</td>
</tr>
<tr>
<td>The compilers interpretation of the computable data entity</td>
<td>A recommendation on how they should be interpreted and used</td>
<td>An applicability guide for the data user supplied by the provider of the data</td>
</tr>
</tbody>
</table>

Each computable data, within a data set on an activity, may have been retrieved differently. Therefore, this information can be supplied for each numerical value.

5.3 Meta data on receiving body

For the impact assessment stage of LCA it is necessary to have information on the environmental interference caused by the free flows (described in section The Concept of Flow) of the technical system.

Figure 4 shows a finite technical system with free flows. Some of these free flows reach another technical system, either as raw material or energy supply, or as products and waste left for waste treatment. The rest of the free flows reach the environmental system as resource extraction, or as emissions or waste. In general, impact assessment methods regards the environmental system as one large "average environment". To perform an impact assessment applying such methods, free flows, summed parameter for parameter, is not sufficient. Additional information must be given in order to identify the local or regional environment. When designing SPINE it was
decided to make it possible either to describe the environment in terms of environmental types and/or in terms of geographical locations. When the receiving body is described in terms of locations, the environmental type must be found from external meta data, for example environmental geographical information systems (GIS) or habitat models.

6 Discussion and Conclusion

The data model has been proven successful: It allows full transparency and reuse capabilities of LCI data. SPINE can be modified to store any meaningful meta data on any computable data and on any system property. SPINE also allows inventories of technical systems to be reused like any originally collected data. The model is also methodologically flexible, supporting data handling for several types of LCI methodology.

There are SPINE implementations today, both large databases holding collections of LCA data and commercially available software applications using SPINE as a storage format [16,17]. There are also commercially available Electronic LCI data communication software based on the same model is also commercially available. It should be stressed that data can be displayed in any way a user desires, for example as graphical flowcharts, as SPOLD [9] forms, or via software directly accessible via the internet. This generality has been achieved both by focusing on data modelling, but also from having chosen the relational database technology for the implementation.

7 Future Development

Having a technically and methodologically functional data model and well-functioning database implementations for LCI data is a good step forward for LCA data handling. However, there are still some important steps to be taken.

In order to efficiently exchange and reuse LCI data, LCI databases and other useful data sources need to be compatible with each other. The work of SPOLD [9] has meant much in terms of a common semantics for LCI data, but there still is no common view of the concepts and the logic of such a format. We recommend that a common view of LCI data, in all aspects, should be done by applying a data modelling approach.

A consequence of what was described in the subsection Meta Data on Receiving Body, SPINE is designed as a module for data on technical systems. Any compatible module modelling the environmental system can be joined with this module. It would be of great interest and importance to the LCA methodology to have good, compatible models of the environmental system. To fully exploit the capabilities of LCI data models, such modelling still needs to be performed.

8 References