Establishment of trade structures for LCI data

A report describing CPM's strategy
for the development of LCI data trade

Raul Carlson, Maria Erixon and Ann-Christin Pålsson, 2000

Summary

According to the decision made by the CPM Board of Directors on 5 February 1998, CPM's database is to be commercialised.

This report suggests that commercialisation be implemented by creating an infrastructure for LCI data trade. This infrastructure will then be made up of a number of databases, owned and managed by trade institutes, research institutes, et al., with the same structure, data management and foundations for pricing. A fully developed infrastructure of this type would solve the real problem: greater good-quality access to LCI data for use for various types of LCA study.

This report describes a number of important areas which are required in order to establish data trade. Of particular importance is the PHASETS model, which describes what is handled, and how. However, this report is not exhaustive in this respect.

The role of CPM in this infrastructure is to provide *acknowledgement* for the people working in a manner accepted by CPM, described in this report, and to allow these people to refer to CPM in their marketing.

CPM is able to commercialise the inspection and acknowledgement, and it can even commercialise its own database, using the same acknowledgement as a basis. The intention behind this document is to provide support for data management and data sales, and it therefore places particular emphasis on data management routines on the one hand and on the inspection which can be carried out by CPM so as to be able to acknowledge the management procedures of organisations on the other.

August 2000

Contents

1 Introduction

Centrum för produktrelaterad miljöanalys (CPM, the Centre for Product-Related Environmental Analysis) started in 1995, and one of its aims was to establish a national LCA database. Its results included a computer documentation (communication) format and an LCI database based on the SPINE concept.

The first CPM project, the “Database Project”, comprised various subprojects which dealt with data collection and data input [1]. These showed that the amount of available data in the various databases within CPM, at companies and a college, was insufficient to satisfy the needs for data within the same networks. It also became clear that many of the companies were in possession of copies of the same information and also had similar needs for information; particularly, of course, the companies working in the same fields.

Given these facts, it has turned out that the most important result from the database project is the data management methodology which is presented in part in this document.

In the early 1990s, the most obvious solution to the problem of barely sufficient access to LCI data was to create a database. Since then, however, it has become more and more clear that the well thought out, efficient production of data is an even more important target if access is to be improved. Data will not exist, either in databases or anywhere else, unless better solutions for producing data are developed.

The CPM SPINE@CPM database was established during the database project, and the Board of Directors at CPM has decided that this database shall be commercialised in order to give everyone access to the data in it in a structured manner. This is the formal background to CPM’s proposal for a commercial form for LCI data management. The objective of this report is to describe CPM’s plans and intentions as regards the sale and exchange of LCI data.

It is proposed that the commercialisation of the database, in accordance with the decision of the Board, be brought about by establishing a trade structure, instead of just selling the LCI data in SPINE@CPM. The reason for this is that the latter option would be extremely short-term. Nowadays, getting hold of data is expensive, and it is often of low, unknown or varying quality. Moreover, the need for specific data, such as for certain industries or products, is very variable. Therefore, pure commercialisation of CPM’s database would, as things stand at present, be merely a short-term solution which barely covers the great and numerous needs that exist.

Establishing an infrastructure in which many parties have financial incentives to contribute towards the development of the system is a more long-term solution for a number of reasons. For example, efforts of this kind are required if we are to be able to apply, long-term, the quality-conscious methodology prepared within CPM’s data activities. The commercial strength of the methodology surrounding CPM’s management of data cannot yet be assessed fully on account of the fact that it is new. Understanding of this methodology must be circulated throughout the entire data management system so that consistent quality improvements are possible, thereby leading to an increase in the accessibility of LCI data. An understanding of this methodology is required for quality inspections to be made cost-effective and for putting trade structures into action, for example.

Another – perhaps more important – reason why it is necessary to concentrate on this long-term solution is the fact that data quality is dependent on inspections being carried out by particularly skilled staff within companies and industries. Allowing a number of people to carry out the inspections, each of whom has the special expertise required to carry out a meaningful inspection, is expected to lead to a general enhancement of the quality of the LCI data.

2 Data networks

This report contains a description of the establishment, initiated by CPM, of trade structures for LCI data. Trade structures may consist of a network of organisations and people who wish to buy, sell or exchange data in an orderly manner. A number of different solutions are conceivable regarding how the trade itself would take place. For instance, the following types of data trade involving SPINE@CPM are currently used at CPM:

- Free access to the database for CPM members if 5 SPINE-formatted and documented data quantities are submitted per company *per annum*. The network comprises 15 members in the form of major Swedish companies.
- Free sales of LCI data via the Internet.
- Free access to the database for people studying for examinations studying LCA or something similar if they submit in SPINE format their entire system studied.

There have also been a number of discussions on data exchange agreements with other colleges and universities, where a longer-term, more continuous exchange may be relevant. A number of ongoing institute-oriented and trade and industry-oriented projects may also result in an enlarged data network as these are based on the methodology produced at CPM.

In data networks, it is necessary to share the same views of data, its form and quality, and so on, partly to be able to share the work collecting data (some parts of a life cycle may be common, e.g. as regards transport, energy production and the production of packaging), but also to be able to divide existing data.

Networks require the following components:

- Rules for network activities
- A common language in the form of terms, terminology and syntax

- A possible communication path

Possible communication path relates to technical solutions. These are not dealt with in this report. However, CPM work is ongoing in this field, e.g. the development of the SPINE communication file .xfr [2] and the LCA data documentation format ISO 14048 [3]. Other parts are described in outline in the text below and in more detail in the next section.

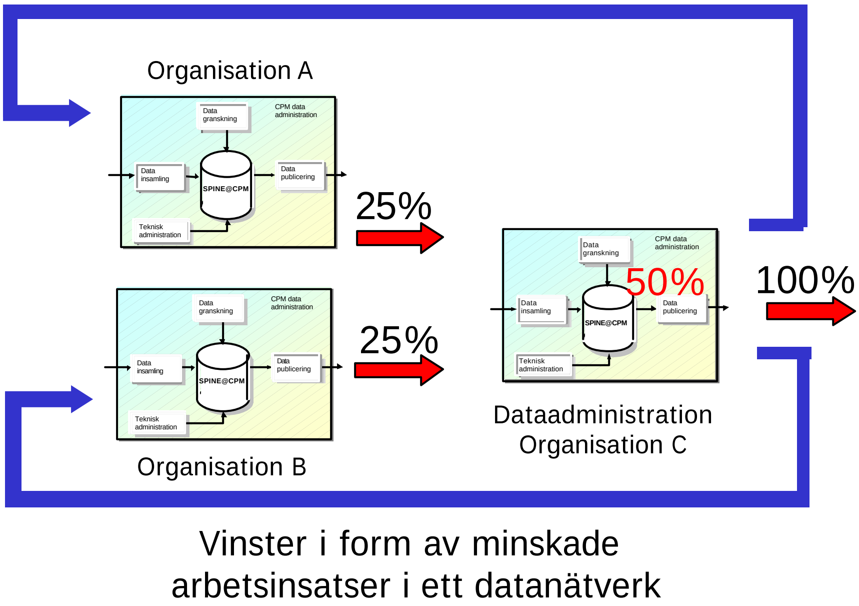

Figure 1 A schematic example of gains in the form of reduced labour input which a computer network can offer organisations with their own data administration.

This report places emphasis on the *trade structure* concept rather than on *data communication*, even though the latter relates more to the technical reality which this involves. Trade, on the other hand, relates to the profitability which can be achieved by all parties when using these network structures. “Trade of LCI data” here means free data exchange between two parties, regulated by means of agreements on quality and price.

Rules for network activities

CPM has frameworks relating to the setup of computer network activities, including quality requirements and instructions on how the administrative organisation should be structured, in particular with respect to responsibility for inspection. CPM has produced an 11-point list describing these rules, and the intention is to give official acknowledgement to anyone selling or otherwise publishing data in accordance with these rules.

A common language

Both sellers and buyers have to have a clear and shared perception of what LCI data is so that they can trade LCI data to a great extent. In contrast to the rules which apply to modestly developed trade, data buyers must also be able quickly to assess the general quality of data management as far as the seller is concerned.

The following are a few important concepts in CPM's data activities:

- Definition of an LCI data quantity based on SPINE
- The PHASETS model, describing data production phases
- The CPM quality concept, where data quality is equivalent to data documentation

Definition of LCI data quantity

The LCI data definition can be further divided into five sub-descriptions: objective and scope of inventory, the system model, instructions for the use of data, inflows and outflows, and methods used to produce inventory data. This definition is based on SPINE and the work done during CPM's first project stage in the database project. See also section 4.

The PHASETS model, describing data management phases

A model has been produced at CPM which covers the various phases of data management, i.e. the production and moving of data from data source to final reporting. This is known as PHASETS (*PHASEs in the design of a model of a Technical System*) and offers a well defined structure and terminology which can be used to discuss and plan data collection, etc. See also section 5.

Data quality synonymous with documentation quality

The quality concept was the first one to be defined within CPM when the initial database project was started. An initial quality definition was agreed upon, but this did not preclude the possibility of it being extended or amended at a later stage. However, the initial definition is still in use. See also section 6.

Quality and cost frameworks

Labelling for traceability

One important qualitative property of data is its traceability, i.e. the options for tracing data back to its point of origin, and the possibility of determining whether data has been managed correctly in relation to what it is alleged to describe. The PHASETS model can be applied in order to classify and label data for the purposes of traceability. This is then described in terms of "visual depth". See also section 7.

Price estimates

To be able to establish trade, joint pricing offers both support and a target for both sellers and buyers. CPM refers once again to the PHASETS model, which here includes various forms for cost allocation and price fixing. See also section 8.

3 Rules for network activities

Any organisation which intends to publish LCI data with good quality defined in accordance with CPM criteria must meet the following 11 points. Each of these points is described in more detail below.

1. Keep data in order
2. Give one person responsibility
3. Ensure that there are sufficient tools
4. Clarify the forms for publication
5. Clarify input procedures
6. Be clear to data users regarding what rules apply for the data being published
7. Clarify the phase(s) at which data management will take place on the basis of the PHASETS model
8. Stand responsible for the inspection of data in respect of the criteria set up by CPM
9. Ensure that the management of the organisation bears ultimate responsibility for the correct management of the data
10. Create an organisation document which describes points 1-9

11. If an organisation is to be permitted to refer to CPM or its quality criteria in respect of data and data management, CPM must have been given permission to check that points 1-10 have been complied with to its satisfaction.

1. It must be simple to look up and identify data in the system. The organisation must have a person-independent system and know all about what data is:
   a. Approved data
   b. Working material, i.e. data collected which has not yet been approved as inspected data
   c. Outline data, i.e. data which will not be inspected

2. One person must be responsibility for ensuring that:
   a. Order is maintained in accordance with (1)
   b. Everyone working with the system has the time and the authorisations to carry out their work correctly
   c. Inspectors and others have sufficient knowledge to be able to carry out their work correctly and follow procedures
   d. Other procedures in this document are followed

3. Sufficient tools for maintaining the system must be available, such as databases, software and manuals.

4. There must be a well defined publication structure:
   a. User groups must be well known, i.e. everyone using the database must be registered
   b. Pricing must be clear and data costs entered in the books. The pricing must be a conscious, documented decision

5. Input procedures must exist:
   a. Prior to each data input situation, how and when data inspection is to be implemented must be

defined, as must who is responsible
b. While work is in progress, each newly collected data quantity will belong to the Working material category (1.b)
c. After inspection, each data quantity will belong either to the Approved category (1.a) or the Working material category (1.b)

6. As regards CPM data management inspection:
   a. Only approved data (1.a) may be connected with CPM data management inspection
   b. Neither working material (1.b) nor outline data (1.c) may be connected with CPM data management inspection

   Note, however, that CPM data management inspection does not determine whether data in accordance with category (1.b) or (1.c) may be communicated or sold.

7. Define activities on the basis of the various data production situations in PHASETS (see the section entitled Model describing data production phases):
   Phase 0: definition of the scale of the selected measurement parameter and its significance
   Phase 1: original production of individual numerical value
   Phase 2: original production of statistical frequency function, based on a number of individual numerical values
   Phase 3: original compilation into a data quantity describing a technical system, in accordance with definition of LCI data quantity (see the section entitled Definition of LCI data quantity).
   Phase 4: original aggregation of data quantities to form a new data quantity (average value formation, LCI and the like)
   Phase 5: moving data between two context environments

8. Data inspection procedures
   Inspection procedures vary depending on the data production situation in which data is to be inspected. According to the above production situations, various aspects of data will be inspected, but as we are referring to LCI data here, we will restrict ourselves to inspecting the quality on the basis of phase 3 or 4.

9. Responsibility for ensuring that the demands of CPM data management inspection are met lies with the management of the organisation.

10. It is the responsibility of each organisation to document how points 1-9 are complied with.

11. The CPM data management inspection deals with:
    a. documents according to point 10
    b. random inspections of approved, inspected data

4 Definition of LCA data quantity

LCA data refers to data which can be used in life cycle analyses. An LCA data quantity always applies to a specific technical system where material and energy are used in some way to perform a desired function, e.g. manufacture of a product. A technical system can also be an aggregation of various technical subsystems of this type, in accordance with the SPINE model [4]. This function gives rise to environmental impact in the form of emissions, waste, etc.

According to CPM, the definition of an LCA data quantity – see figure 2 below – includes descriptions of: objective and scope of the life cycle inventory, the model of the technical system, recommendations for the use of data, inflows and outflows, and methods used to produce inventory data and important information for assessing the usefulness of data.

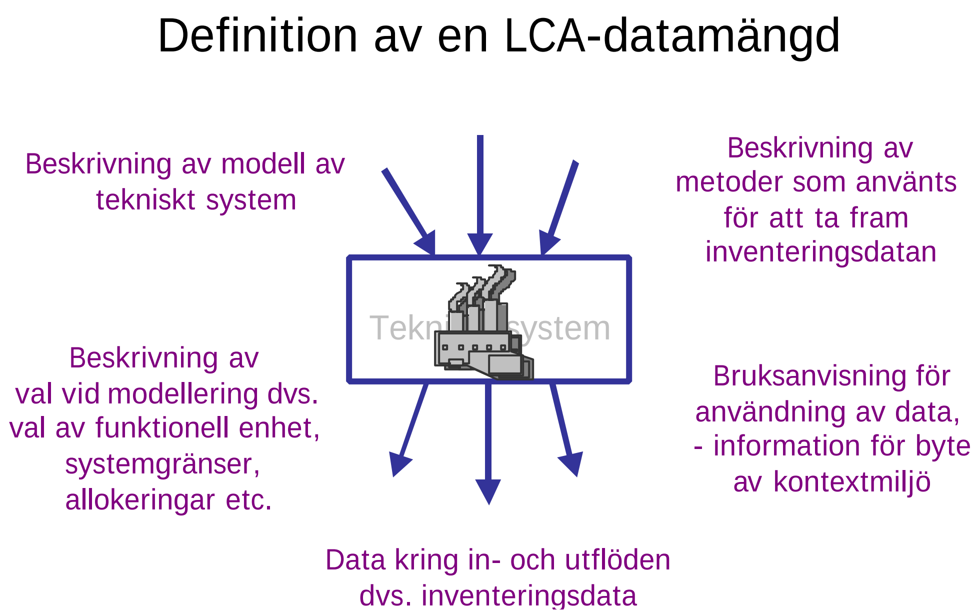

Figure 2 This illustration describes CPM's definition of an LCA data quantity.

If a data quantity is to be identifiable and interpretable in respect of relevance for use in a specific LCA study, the technical system and inflows and outflows must be sufficiently well described [5]. What is meant by "sufficiently well" depends on the knowledge level and prior understanding of the person who will be using the data quantity in order to carry out an LCA study or to inspect a data quantity or a study which has already been carried out. A clarification of what constitutes "sufficiently well" documented data is included in the report "Introduction and guide to LCA data documentation using CPM documentation criteria and the SPINE format" [6].

5 Model describing phases of data production

"Data production" is a very general expression which means completely different things depending on the type of data to be produced. The measuring of the flow of flue gas from a chimney assumes specific measuring equipment, a specific measuring methodology, and an awareness of other technical criteria when the measurement is taken, while the production of a specific collection of average value data for the manufacture of steel, for example, presupposes a knowledge of the current steel market, a knowledge of comparability in respect of any allocations when producing the various data quantities, the use of appropriate, correct statistical methods, a clear objective regarding which data is to be produced, and so on.

Generally applicable methodology for LCI data management necessarily includes the production of all the different kinds of data used in LCA. In various contexts, it will also be necessary to be able to refer to a specific data production situation. For example, this applies when designing documentation criteria in respect of data, where CPM's requirements to date have been formulated for a complete LCI data quantity as it must necessarily be documented in a general national LCI database. However, to produce a complete, compiled LCI data quantity, it is necessary to first have worked through a number of earlier procedures.

A reference model containing six levels, or *phases*, for the production of data has been produced by CPM's data activities: see figure 3. This model is designed to support the construction of concepts so that it is possible to refer to a specific type of data production. Each of the six phases in the model describes a specific data production procedure which is used when an LCI data quantity is produced.

The six different phases of the model are represented as a 6-phase model where an LCI data quantity describing a technical system is compiled from data from an underlying phase (from the production of an individual numerical value) and finally reported or published in phase 5 (data moved between context environments).

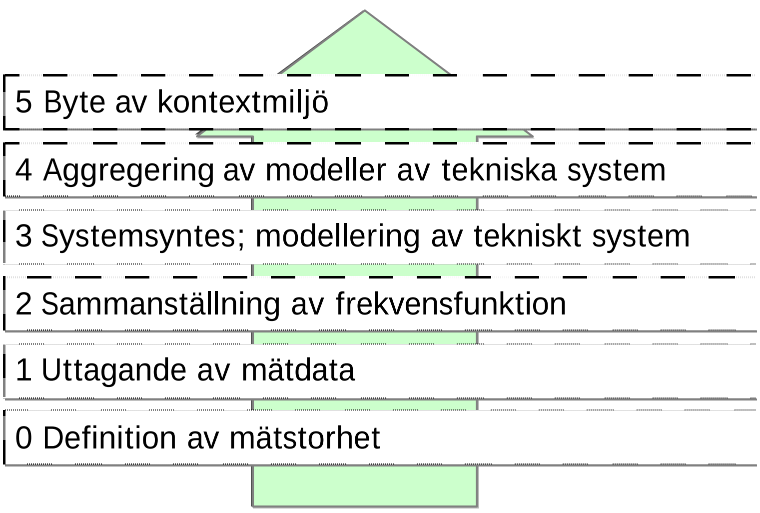

Figure 3 PHASETS Model describing phases of data production

The content of each of these phases is described in more detail below. A further description of the model can be found in the article "Industrial Environmental Information Management for Technical Systems" [7].

Phase 0: Definition of scale of selected measurement parameter

Phase 0 is the foundation for subsequent PHASETS phases. A clear understanding of the scale of the measurement parameter is required for further communication of information to be meaningful. When setting up the measurement system, it is necessary to ask: what do we want to check, so what shall we measure, and why? It may also be necessary to take into account mutual relationships between measurements in the system, e.g. can air emissions be dependent on the production volume?

Phase 1: Original production of individual numerical value

In LCI data quantities, each individual numerical value, such as the size of a flow or length of a section, has been produced by means of one or more measurements, looked up in one or more publications (printed or electronic) or been estimated from experience.

Each of these numerical values has a specific history:

- A measured numerical value has been measured at a certain time, under certain operating conditions, in a specific position in the technical system.
- A numerical value retrieved from a publication has a specific reference and specific documentation which describes the criteria for the value.
- A numerical value estimated from experience is estimated using certain *specific* experiences as reference points.

For an individual numerical value to be regarded as known, its history must be described on the basis of the parameters which are of relevance when data is to be used to compile an LCI data quantity.

Phase 2: Original production of statistical frequency function, based on a number of individual numerical values

With the current LCI data situation before us, it may seem excessive to have a specific phase for the production of a statistical frequency function, but there are two reasons for this:

- In phase 1, only the production of individual numerical values is dealt with. In most cases, even a fairly rough estimate is based on an appraisal of a number of individual numerical values, such as: "It absolutely cannot be less than X and it is physically impossible for it to be more than Y, which means that I will assume that an appropriate typical value is Z."
  This reasoning comprises two individual numerical values, X and Y, and a subjective estimate of the frequency function Z. X and Y are individual values, produced in phase 1, while the estimate of the type value is part of phase 2.

- In an actual measurement situation, a number of values are as a rule measured in accordance with phase 1, perhaps several per second or one or more over a year. *Different* statistical methods, filtration methods, estimates and extrapolations are required in order to create a meaningful frequency function from each of these two types of data quantity.

For a frequency function to be regarded as known, the methods used to produce it must be described, irrespective of whether these are purely subjective or strictly mathematical statistics.

Phase 3: Original compilation into a data quantity describing a technical system, in accordance with definition of LCI data quantity

Phases 1 and 2 describe typically numerical values for inflow and outflow sizes to a technical system, and according to the definition – see figure 2 – an LCI data quantity is meaningful only if the inflows and outflows are regulated in relation to a well defined technical system.

It is common for a manufacturing plant, for example, to supply more than one product, which means that the overall production capacity, such as machinery, support functions and premises, is shared by the various products. It is also common for the production capacity to be distributed *differently* to the various products, which leads to what would be known in an LCA context as an “allocation problem”.

To put this differently, this means that the actual plant is divided up into a number of “pretend” processes, each of which manufactures its own products from the range manufactured by the plant.

In this case, identifying the inflows and outflows of the individual products from the plant is often a relatively complex problem: first of all, it assumes good models of the “pretend” processes so that they describe the actual case as well as possible. This is often trivial. But this also demands that the flows measured and compiled at the plant, according to phases 1-2, are recounted so that they are transferred to the inflows and outflows and standardised in relation to each and every one of the “pretend” processes.

The models of different types which are used to solve the “allocation problem” are seldom purely physical descriptions, but they often demand a number of less well supported assumptions, estimates and criteria. This fact is independent of whether the allocation problem is dealt with on the desk of an LCA analyst, with only a few facts on which to base the model, or whether this is done within the plant. The difference between these two cases is only the fact that the latter could conceivably have more information available on the actual conditions.

Nor is there any difference if the allocation problem relates to a small technical system with a few products or a large technical system including a number of plants, distribution systems and energy production, for example. However, the latter makes the problem more complex and makes both modelling and documentation more difficult.

For a technical system to be regarded as known, it is very important that the solutions to allocation problems are described well.

LCI data quantities are described as known by a number of inflows and outflows, among other things. Usually these are by no means all inflows and outflows into and out of an actual manufacturing plant, but only those which have been selected for some reason as being particularly important from the point of view of the environment when the plant, or a "pretend" process from this, is modelled together with an LCI data quantity.

Which flows are considered to be significant depends on which environmental model was used for the compilation. This may be a well formulated model, such as an LVA environmental assessment method, or it may be more vaguely formulated in economic terms, for example, such as "It is these flows which the management considers to be so important that it has chosen to invest in measuring equipment to measure them".

Irrespective of the reasons for why a certain set of inflows and outflows has been selected, it is important for the person who will be using the data quantity to be familiar with these reasons. As a rule, it is far too difficult for an LCA analyst to determine whether specific flows have been "forgotten", and if so, why this is so. This is best determined by the person who compiles the LCI data quantity, and therefore it is also best described by this person too.

Phase 4: Original aggregation of data quantities to form a new data quantity (LCI, average value formation, etc.)

LCI studies

LCI studies begin by defining an objective and a scope for the study in question. During this stage, an initial model is created relating to how the product flow should be followed backwards towards the extraction of raw materials and forwards towards waste management or recycling. All along the product flow, it is usual for a number of "branches" to fork off, either taking into account or ignoring the allocation problems.

This modelling stage often results in a description flowchart which contains a number of "small" technical systems which are filled with data from "pretend" processes (as described in phase 3 above) and transports, for example.

The result of the inventory is presented as an overall environmental profile for the entire large technical system represented by the product life cycle studied. In SPINE, a "large" compiled technical system of this kind is known as an *aggregated activity.*

For an aggregated activity to be regarded as known, it is important that the modelling of the product flow, i.e. the decisions made when defining the scope and objective of the study, is described well. However, it is also important for people to remain familiar with the mathematical methods used to calculate the environmental profile, as there are a number of different methods which can be used for calculation which all can give different results in different situations.

Average value data

One common way in which industries are able to manage data confidentiality and distribute LCI data at the same time is to compile various types of average values for the production of trade products. This can be done both in a “cradle to gate” phase and at plant level.

Average value formation for various types of technical system is also done on a general level in order to form what is known as “general” data for transports, the production of energy products, various common raw materials, etc.

When the average values are formed by a number of technical systems, great analogies can be drawn with the aggregated activities belonging to LCI: a mathematical and/or statistical method is permitted to act on a number of technical systems in order to form a new aggregate, the average value system.

On the whole, this applies to the same documentation criteria for average value data as for data quantities aggregated in LCI.

Phase 5: Moving data between two context environments

When data is retrieved between various phases of the PHASETS model (phases 0-4), or when data is exchanged between departments within a company or between different companies, or when data is moved between different LCA databases, for example, this can be achieved, from the point of view of interpretation, in three different ways:

1. Data is interpreted in the same way in the receiving phase or environment as it was in the phase or environment which issued it.
2. Data is interpreted differently in the various environments, but this fact is ignored, or any reinterpretations are made verbally between the information provider and the data recipient and not added to the documentation on the data quantity.
3. Data is interpreted differently in the various environments and reinterpretation is carried out jointly by the information provider and the data recipient using the best possible translation, and this interpretation is transferred to the data quantity so that the data can be interpreted correctly in the recipient environment.

The consequences of alternatives 1 and 3 are the same: data is interpreted correctly in the new environment and can be used directly by anyone in the recipient environment. The consequences of 2, on the other hand, are that data will in any case be difficult to interpret or will be interpreted incorrectly in the recipient environment, irrespective of how well the data has been documented by the information provider.

The natural thing is for data to be expressly reinterpreted in the recipient environment. This is true in particular when data is to be published for a wider audience, which is the case with published LCA reports, the type III environment marking system and public LCA databases, for example.

Reinterpretation can be carried out in many different ways, depending on who is doing the interpreting and the objective and aim of the reinterpretation. Reinterpretation also always involves a risk of distortion of the original content and the introduction of errors. Therefore, reinterpretation of text, for example, must always be documented.

For a data quantity published by anyone other than the original author of the data to be regarded as known, it is important for all forms of reinterpretation, such as reformatting, filtration, etc., to be well described.

6 Data quality synonymous with documentation quality

CPM’s definition of data quality is based on three quality aspects: *reliability*, *accessibility* and *relevance*: see figure 4. All three parts are equally important to be able to investigate and assess whether the quality of a data quantity is suitable for a specific study. These terms are described in detail in the article *Fundamental Aspects of Data Quality* [8]

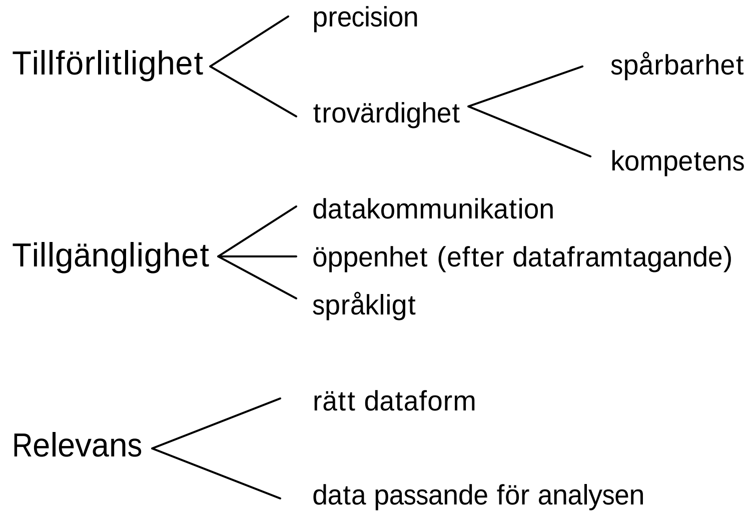

Figure 4 Definition of data quality according to CPM.
© Carlson and Pålsson, 1999

7 Proposal for the marking of traceability for LCI data

According to the PHASETS model described in “Model describing phases of data production”, the production of data can be divided up into six phases, and the same phases can be used to describe the very important qualitative property *traceability* in the case of LCI data.

For data at a specific phase to be traceable to its source, the production of data within each phase up to and including the present one must be documented correctly: if, for example, an LCI study is implemented and documented (phase 4), documentation of the way in which each of the technical subsystems included has been compiled (phase 3), documentation of each of the frequency functions included in the study for each of the flows (phase 2), documentation of the production of each of the numerical values used to build up the frequency functions (phase 1) and documentation of the scale of selected measurement parameters and related information are also required. Also required is the documentation of each context change between phases (phase 5). If, for example, phase 2 is left out, phase 1 cannot be included either, which means that in this case data is traceable only to phase 3.

Nowadays, it is most common for data to be traceable only to phase 4, so the data which is included in an LCI study is not traceable even with respect to how the technical subsystems included in the system have been modelled.

Data can thus be marked with a traceability figure according to the PHASETS model, which represents how far it is possible to trace a data quantity back to source. This can also be called “visual depth”.

8 Models for estimating the price of data

Price of data

Users of data should pay a price (P) which is greater than or equal to the cost (K/N = dK) of providing them with data:

$$(AN + FP + RD) = K$$

$$(AN + FP + RD)/N = dK$$

$$dK \leq P$$

AN = Cost of obtaining data
FP = Cost of administration/publication of data
RD = Cost of further development for administration/publication of data
N = Number of data users
K = Cost per data quantity
dK = Cost per data quantity for one user
P = Price per data quantity for one user

If P = dK, this means that the database owner will not make a profit on the data published.

RD: Model for further development costs

LCI data management is a new area, which is why it is likely that all profits made in the next few years will be used for further development.

Standardisation, the production of tools, the improvement of the statistical qualities of data and clarification of the LCI system description will be developed considerably over the next five to ten years. The costs for this development will be borne jointly by all parties participating in the development/use of LCI data and data systems.

It is difficult to provide a model for these costs in general, but they could possibly be estimated by each individual data owner on a yearly basis, for example: in this instance, it may be reasonable to set these costs at the same level as the annual budget for development projects with respect to the improvement of data management.

It is also reasonable to assign the training of data input staff to these costs as well at present, as at the moment special training is required so that these people can learn how to format data and document it correctly for the system.

FP: Model for administration/publication costs

The costs involved in the administration/publication of data (FP) are as follows:

- Costs for technical subsystems, such as licences for databases, software and servers (TK), and
- Costs for data inspection. See separate model on costs for data inspection (GK).

This means that the costs involved in the administration/publication of data can be expressed as follows:

FP = TK + GK

AN: Model for the cost of obtaining data

Of course, it costs different amounts to produce different kinds of data, depending on the situation and the data type: inputting a single digit has its costs, and producing a “cradle to gate” data quantity involves different costs. Moreover, costs are allocated differently in different situations, depending on the driving forces behind the production of the data. Therefore, any attempts to estimate costs for data production need to be supported by some kind of structure, and it is a natural step to use the PHASETS model described above.

To be sure, costs can vary widely within each phase, depending on tools, scope, ambition, etc., but the complexity of the management and the methods which operate on data are similar in the various phases.

In the long run, underestimating the actual costs will lead to the stopping of production of new data as there will then be a risk that the costs will not be covered by the price. Therefore, it is important that each organisation producing data, at each phase, is aware that there is nothing to be gained by underestimating costs. It is probably a good idea to err on the side of caution and overestimate costs slightly.

The costs at each phase are a total of three subcosts:

- The cost of interpreting data prior to processing in accordance with phase 5: see the section describing the PHASETS model:
  fN5
- The cost of processing data in accordance with the current phase n:
  Nn
- The cost of providing an interpretable explanation after data processing, i.e. reporting/documentation:
  eN5

AN = fN5 + Nn + eN5

Phase 1: Original production of individual numerical value

An individual numerical value can be obtained from a number of different kinds of contexts: from literature, as a result of a measurement initiated specifically to produce a single digit, from a logging system which automatically measures with a certain level of frequency, by asking someone, etc.

The cost of the individual value does, of course, depend greatly on how data is produced. Creating, maintaining and using an automatic measuring system does, of course, cost a lot more than just calling someone and asking, but there are also often far more reasons as to why an automatic measuring system has been set up than there are reasons to just call someone and ask about a numerical value.

For example, it is possible that measurement will take place in order to control a process, and in this case all costs are allocated to the measuring system for the actual process. The costs of retrieving a value from this system for other reasons are then likely to be very close to the cost of a telephone call. There are also many other instances where the entire cost of data can be allocated to the process, such as when the measurement is a legal requirement to be able to run the process at all, or when economic or other administrative control is based on the values measured.

Thus the cost of an individual numerical value is often negligible from the point of view of LCI, but it may be of significance for individual values if there is no incentive in other systems for producing such data.

When the cost of individual values is to be produced, reasonable allocations must be made, but actual costs should not be concealed. For example, actual costs include working time for reading the value and the time needed to sufficiently document the measurement.

Phase 2: Original production of statistical frequency function

Statistical frequency functions are formed either as strictly mathematical statistics, or, less strictly, from a number of individual numerical values. Whether the frequency function is produced using strict methods or from more subjective assessments, the results are *estimates of the actual value*.

The cost of producing an estimate is related to the cost of the working time and tools used.

Costs are allocated to the various objectives for which the frequency function is produced.

Phase 3: Original compilation to form a data quantity describing a technical system

Frequency functions, as per phase 2, may be compiled to form a data quantity which describes a technical system for many different reasons: for environmental investigation, for environment reporting to the county administrative board, or for aiming the environmental impact of a product towards the product target during production.

Usually, each compilation of this kind requires some kind of modelling. This modelling includes a number of important parts:

- Definition of technical system, including:
  - System function or product, and process parts included
  - Allocation, where flows to and from an actual or modelled manufacturing plant which have an impact on the environment, for example, are allocated to the relevant product(s) or other functions.
  - Remodelling or ensuring that all the various measurement values which the frequency functions describe belong to the technical system being described.
- Consistent description of the background to the choice of parameters which are considered to describe the inflows and outflows of the technical system.
- Other technical properties, as specified in CPM's SPINE documentation criteria (reference manual).

Normally, this work is demanding, both intellectually and in terms of resources, and it is best if staff with a good knowledge of processes implement this internally at the plants.

If the compilation is carried out in order to steer production in respect of environmental properties, e.g. with respect to environmental targets for plants or products, the cost is allocated to production. If reporting is carried out for product communication, the cost is allocated to production information or marketing and thus distributed over the entire cost of the product.
If special reporting is carried out at the request of a trade institute, for example, the cost is allocated to special costs for trade administration.

Normally, the same reports would be compiled for all of these objectives, which is why the often significant resources demanded by this work may be relatively modest for each objective.

If the work is done by an LCA analyst or other expert "at his desk", there are usually fewer objectives, but in this instance it is also reasonable to assume that the work is largely intellectual and demands fewer resources in other respects. However, the cost of intellectual work must not be underestimated as it has to be carried out by one or more experts with a special knowledge of processes.

Phase 4: Original aggregation of data quantities to form a new aggregated data quantity

As in phase 3, special process knowledge is required in order to be able to create different types of technical system aggregates.

A thorough investigation of the entire life cycle of a product or service – or of that part of the life cycle covered – is required for an LCI study. A series of well considered decisions on system modelling and data quality are made, and most of these decisions can be made in a number of different ways. The result of the aggregation is entirely dependent on these various decisions.

When compiling average values, such as trade averages and annual averages, another type of decision has to be made on system modelling and data quality. In the case of trade averages, a number of – often very different – production plants should be compared and a statistical average estimated. This is often very difficult and expensive, depending on the ambition level. Calculations of annual averages for individual plants is slightly less complex, and these can be more or less difficult to implement depending on how these are made.

In general, the cost is described by the time taken for the work, plus the cost of the tools needed to do it. As usual, the cost is allocated to various parties.

Phase 5: Moving data between two context environments

At present, this phase is the most demanding part of an LCA study in terms of cost. Producing data from various sources, interpreting it and reformatting it is difficult and expensive. Consequently, moving data between context environments is a great source of errors and uncertainty: there are seldom resources available for searching for the correct data for all the technical subsystems included in a study, and it is often difficult or far too demanding in terms of time to correctly interpret the data produced.

Data is moved between context environments between each phase 1 to 4, from measurement logs to statistical data management (1 to 2), from statistics to reporting (2 to 3) and from reports to aggregation to LCI studies, or for the compilation of trade averages (3 to 4), for example.

The correct transfer of data includes the data issuer and "customer" together going through and interpreting the data quantity transferred in order to ensure that the data has been interpreted – and will therefore be used – correctly. In the case of continuing "business relations", it may be sufficient to implement this interpretation once for several "purchases", and both parties will then guarantee that they will keep to the agreed interpretation. In the case of one-off "business relations", the interpretation work can be an important cost-enhancing part of the deal.

It is for this reason that the widespread trade of data demands a clear system of rules which can be interpreted in the same way by everyone connecting to this system: the interpretation is then independent of each individual deal and can instead be done once and for all. This means a reduction in the *transaction cost* for data.

The cost of switching contexts is thus a cost which must be borne by both seller and buyer. The seller's cost involves making data interpretable for the customer, and the buyer's cost involves interpreting data. The interpretation includes metadata, formatting, text interpretation and sometimes even remodelling.

If the seller is able to create a practical method for facilitating interpretation on the part of the buyer, the costs for the development and utilisation of the method can be allocated to all the customers to whom data is sold. The buyer can allocate costs to the various objectives for which the data is to be used.

GK: Model for data inspection costs

If the inspection procedures are formulated incorrectly, the cost of this element can be enormous. CPM's inspection procedures are formulated to keep down the costs at each stage by ensuring that it is the structure and interpretability of the documentation that is inspected. Inspecting whether or not numerical values are correct is not done at CPM, as this inspection takes place at phase 5 and data usually reaches CPM after phase 4.

One prerequisite for bringing down the costs of data inspection at CPM is for the people who submit data to have a satisfactory understanding of how data is inspected. The same applies to all different types of LCI database: to keep inspection costs down, the data must be input to the database in the best way possible. Over time, it should only be necessary to carry out random inspections in order to monitor the quality and make sure it has not deteriorated.

The costs of quality inspection are allocated to the various users of the data system.

TK: Model for technology costs

Following the CPM documentation criteria or CPM data management procedures requires no special technology. It is possible to build up a database on paper, in various kinds of format-compatible word processing programs and spreadsheet programs, or in advanced database systems which have been made SPINE-compatible in various ways.

The buyer can allocate costs to the various objectives for which the data is to be used.

N: Model for number of data users

The number of data users is a problem when it comes to estimating, as during pricing it is difficult to determine how many users will want to retrieve a data quantity from a database. A few rules of thumb can be compiled as a foundation for estimates by making clear the audience for which the database is intended.

A few important types:

- **National database**: Intended for a large but limited data market. Try to estimate the number of possible data users in this group, take care to establish a relationship with the data users so that the number can be updated over time.
- **International database**: Intended for an unlimited data market. Very difficult to estimate.
- **Industry data**: Estimate the number of users according to how the industry forms part of the data requirements of other industries. This is done by looking at how the products from this industry can be included in general production. Industries which supply basic materials such as paper, wood, concrete, steel and other metals have a very large number of data users, while industries for more specific products have a less general value and, accordingly, fewer data users. Of course, these estimates are difficult, and they should be updated over time depending on experience.
- **Company data**: It is reasonable to assume that companies mainly provide their customers with data. Therefore, the number of data users can be found in the customer list.
- **Internal organisation database:** In the case of internal databases at institutes or companies, it is relatively easy to estimate the number of users, in particular for internal invoicing and the like.

As the number of users defines the data price in accordance with the model shown in this section, it is important to constantly update the number of data users so that the estimates implemented in the beginning can be corrected on the basis of the actual results.

9 Roles and tasks of CPM in the trade structure

CPM is able to assume a number of important roles in a trade structure of this kind. One important aspect here, of course, is the future of CPM as a skills centre at Chalmers. Further development of network activities relating to the distribution of responsibilities and tasks, for example, will be put forward as internal projects within CPM, with external parties being invited to take part.

An operational data market will not only be dependent on a rules system, as described in this report, but also by the fact that there is an independent party with recognised skills and a high level of integrity which may be entrusted with the task of checking that the rules system is being observed, and this may constitute a natural platform for co-ordination and further development. At present, CPM meets the demands for skills and integrity thanks to its overall competence and its close links with Chalmers. It is assumed that CPM also has the support and trust of the industry for becoming an organisation of this kind.

CPM data management inspection

During the CPM database project, database work at Chalmers has taken into account methodology in particular; especially methodology in respect of data management and data documentation, data inspection, technical administration and data acquisition, and data publication. In most of these areas, there is still a lot of research and development work to do, but the results are already forming a good structure for the building of various data administration systems of the same type.

It is intended for CPM:

- to assist other institutes and companies in the use of this structure in the manner described in this report.

- to inspect these other data administrators/publishers and give them temporary CPM acknowledgement. This acknowledgement may be withdrawn in the event of misuse, and it is also intended to implement further potential development over additional phases in future.

- to also train data buyers in the same methodology so that they are able on the one hand to administer their own data, and on the other in order to give them an understanding of what data administration with acknowledgement by CPM involves.

10 Development opportunities in the near future

A knowledge of LCI data collection, data documentation and associated quality aspects is important if a data management system is to work well. Financial backers of industrial research and technical development should therefore support the construction of networks, using the components designed at CPM, in order to implement, cheaply and efficiently, a trade structure for LCI data which works well.

References

1. Establishment of CPM's LCA database, Carlson R, Pålsson A-C, CPM Report 1998:3, Chalmers tekniska högskola, Gothenburg
2. Facilitating Data Exchange between LCA Software involving the Data Documentation System SPINE, CPM et al., CPM Report, Chalmers tekniska högskola, sent for printing in August 2000
3. ISO/CD 14048 LCA Data Documentation Format
4. SPINE, A Relation Database Structure for Life Cycle Assessment, Carlson R. et al., 1995, Chalmers Industriteknik, Chalmers tekniska högskola, Institutet för vatten- och luftvårdsforskning, Gothenburg
5. Krav på datakvalitet CPM:s databas 1997, CPM-rapport 1:1997
6. Introduction and guide to LCA data documentation using the CPM documentation criteria and the SPINE format, Pålsson A-C, CPM Report 1999:1, Chalmers tekniska högskola, Gothenburg
7. Integration of Acquisition of LCI Data within the Environmental Management System, Carlson R, Pålsson A-C, Submitted to Journal of Cleaner Production in December 1999
8. Fundamental Aspects of Data Quality, Carlson R, Pålsson A-C, in progress, CPM, Chalmers tekniska högskola, Gothenburg