Maintaining Data Quality within Industrial Environmental Information Systems

Ann-Christin Pålsson and Raul Carlson[1]

Abstrakt

Ein System ist entwickelt worden, das auf dem SPINE-Datenmodell basiert. Das System ist modularisiert, einschließlich der SPINE-Datenbanken, der besonders geentworfenen Daten, die Software anfassen, der dynamischen Web-Seiten für Datenbankpublikation und, die des wichtigsten Teils, eine Methodenlehre für die Datenbehandlung und –kontrolle.

Es wurde versucht, um eine allgemeine Annäherung zu erreichen, die jede quantitative korporative Umweltinformation anfassen kann. Das Ziel ist, mit dem ankommenden Standard innerhalb der ISO 14 000 Serie einzuwilligen. Dieses Papier schränkt korporative Umweltinformation ein auf Sein Informationen, die die Abhängigkeit zwischen den korporativen Aktivitäten und der natürlichen Umwelt System beschreibt.

Es gibt die grundlegende Undurchführbarkeit, die mit Einschätzung der Datenqualität der Umweltinformation betreffend ist beide Modelle der technischen Systeme und der numerischen Korrektheit der quantitativen Parameter dazugehörig ist. Diese Undurchführbarkeit wird in diesem Papier, sowie eine Annäherung behandelt, die auf methodologischen Vereinbarungen und Unterlagen basiert ist.

Um Befolgung der Datenqualitätsvereinbarung sicherzustellen, ist eine Kontrolle Funktion mit dem System umfaßt worden.

Damit das System zur Funktion als beabsichtigte, muß das System mit den vorhandenen Informationssystemen integriert werden und die organisatorischen Programme muß ausgedehnt werden, um Verantwortlichkeiten für das Umweltinformationsystem auch zu umfassen.

Abstract

A system has been developed, which is based on the SPINE data model. The system is modularised, including SPINE databases, specially designed data handling software, dynamic web pages for database publication and, which is the most important part, a data handling and review methodology.

It was attempted to reach a general approach, which can handle any quantitative corporate environmental information. The aim is to comply with the upcoming standard

---

[1] Centre for Environmental Assessment of Product and Material Systems (CPM), Technical Environmental Planning, Chalmers University of Technology, S-412 96 Göteborg,
e-mail: acp@vsect.chalmers.se, raul@vsect.chalmers.se, Internet: http://www.cpm.chalmers.se, http://deville.tep.chalmers.se

within the ISO 14 000 series. This paper restricts corporate environmental information to being information that describes the interaction between the corporate activities and the natural environmental system.
There is fundamental unfeasibility associated with assessment of data quality of environmental information regarding both models of technical systems and numerical correctness of quantitative parameters. This unfeasibility is discussed in this paper, as well as an approach to a data quality agreement based on methodological agreements and documentation.
To ensure compliance with the data quality agreement, a review function has been included with the system.
In order for the system to function as intended it needs to be integrated with the existing information systems and the organisational routines needs to be extended to also include responsibilities for the environmental information system.

1. Introduction

This paper describes the Swedish national LCA-database system (SPINE@CPM) developed within CPM with special regards to data quality maintenance. The system is based on the SPINE data model and relational database structure (Carlson et al 1995), (Carlson et al 1998). The CPM group has attempted to reach a general approach, which can handle any quantitative corporate environmental information. The reason for this is of course the upcoming standard for LCA, which is part of the general standard for environmental management system (ISO 14 000 series).

The quality maintenance approach handles the issue of LCA data quality similarly to any environmental information system holding quantitative data on corporate activities, regardless of whether the data is to be used for LCA or any other decision support method. Regardless of proposals for management and indication of LCA data quality, e.g. (Weidema 1996), most quality aspects of LCA-data are general and need not be expressed in terms of LCA methodology. For data to be useful in LCA, however, there are some specific data quality aspects to consider e.g. choice of relevant flows (de Smet/Stalmans 1996), not covered within the scope of this paper.

Typical information handled within LCA is data sets on production units. Such data sets are combined into a flow model of a delimited technical system, to analyse and assess the environmental impact of a material, product or service throughout its life cycle. The assessment generally include raw material extraction, processing, transportation, manufacturing, distribution, use, reuse, maintenance, recycling and waste treatment (figure 1).

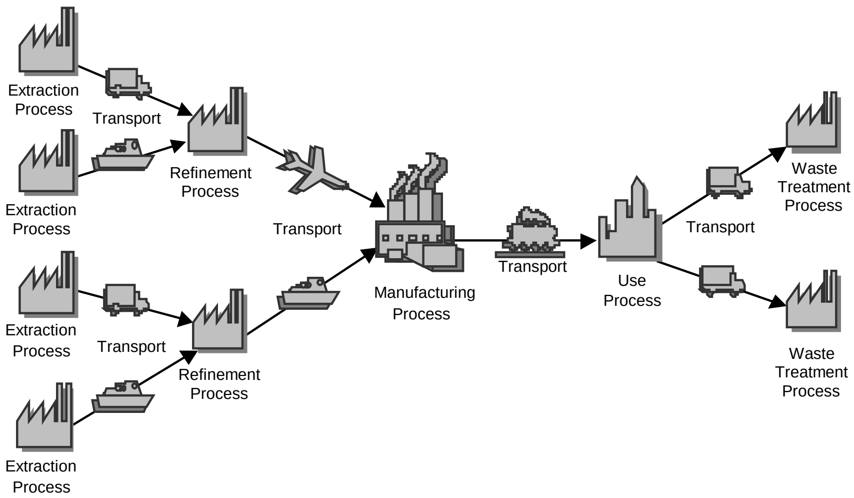

Figure 1

As outlined in figure 1, an LCA generally comprise information regarding different types of technology and different industrial sectors. Considering the possibility that data and information may be communicated between different actors in different sectors, the handling of the information at any site, within any organisation may be simplified if the data is interpreted general and equal, regardless of technology or industrial sector. This also implies that the same data quality routines can be applied regardless of technology or industrial sector.

2. Environmental information systems

Any corporate environmental information system is established in order to improve the possibilities to efficiently control the environmental impact from the corporate activities, with the means of environmental management systems. So is true also for the LCA database and environmental information system SPINE@CPM, which therefore is partly industrially financed.

Examples of areas to control in regards of environmental impact include operation and maintenance of machinery and production plants, product design and usage of products produced and different strategic actions towards markets within which the corporation partake (Netherwood 1996).

2.1 Environmental information on corporate activities

In this paper we restrict corporate environmental information to being information that describes the interaction between the corporate activities and the natural environmental system. It is considered that the activities physically are performed by technical systems and that the physical interaction between the natural environment and the activities is performed by physical flows to and from technical systems (figure 2). Most of the interactions are not direct, but causal via other activities and technical systems. The interactions are described as use of resources and energyware, and as emissions and waste generation associated with the activities of the technical systems.

A corporate activity may for example be the production of a telephone set. The technical system then is the production line within the plant, which enables and performs the production of the telephone sets. The interaction with the natural environmental system occurs both directly through emissions from the plant and causally via the suppliers of components and energyware and via the customers of the telephone sets and via the waste managers, for example.

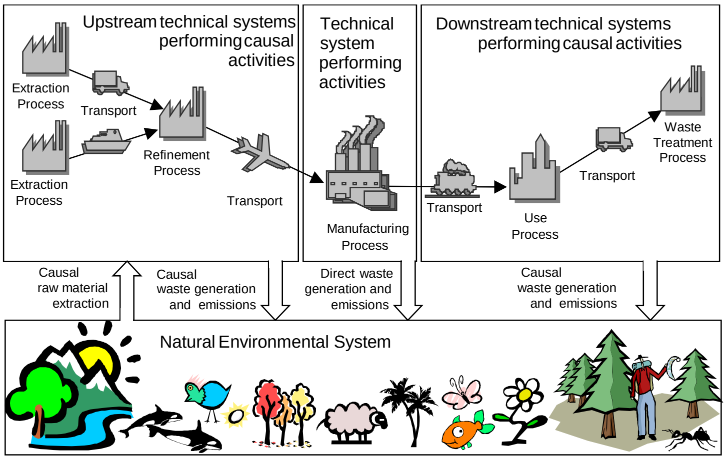

Figure 2

In order to interpret and environmentally assess the corporate activities, the corporate environmental information system also needs to include data on for example the natural environment (Saur 1997). Regardless of the actual importance of such information, it will not be discussed within the scope of this paper.

2.2 Description of technical system

A description of a technical system or subsystem is a model of the system. We will therefore refer to descriptions of systems as models of systems. We will also mean that models of technical systems are manifested as descriptions of technical systems, in terms of natural language, sketches and languages specifically designed for descriptions of technical systems.

Within the Swedish competence centre CPM, a model of a technical system, from an LCA perspective, is defined as (Pålsson 1997):

- a model and a description of its included technical subsystems which defines the system's technical boundary
- a model of its inflows and outflows (matter and energyware), normalised to a system function
- a description of one or more system functions, i.e. a functional unit or functional flow such as products or services supplied by the system

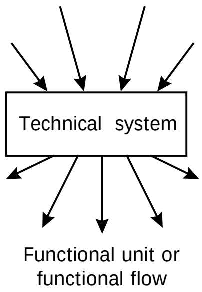

Figure 3

In the example of the production of a telephone set, the included technical subsystems are a description of all operational stages along the production line, including also allocated shares of common production facilities and other support functions. Inflows are described by, for example, data on use of components, energyware, ancillary materials and inflows to common facilities. Outflows are described by data on the production of telephone sets and, for example, data on waste generation and emission outlets from the production and from common facilities and functions. The system function of course is described as the production of telephone sets.

3. Levels of environmental control

Different types of decisions at different levels of environmental control are based on environmental information from within a plant and are combined with other types of information regarding the surrounding systems, both regarding the technical and the natural environmental systems. STORA, the Swedish based multinational forest, pulp and paper corporation suggests the following three levels of control (figure 4)

(Bresky 1998).

Figure 4

Level 1. *Local control between gates of production plant.* To control the environmental impact from the local technical system, analyses of air and water emissions are combined with information of the local environmental sensitivity.

Level 2. *Extended local control.* Extends the local technical system to also include environmental aspects of technical systems controlled by business partners. This type of local environmental control is generally integrated with the environmental management system. This management system often controls and manages the environmental aspects of the choice of suppliers, transports to and from the plant and waste management. This level of control is enabled by combining information of the plant's local environmental information with logistic information and with the local environmental information of the suppliers and waste managers.

Level 3. *Life cycle approach*. The responsibility and the domain of control of an environmental management system can be extended by the application of LCA methodology. Such an extension implies the inclusion of all relevant technical subsystem that can be accounted for in the manufacturing from raw material extraction to the final disposal of the product. This is done to avoid environmental sub-optimisation, such as e.g. the use of scarce resources or environmentally unfavourable waste management procedures for the final disposal of products.

Environmental information on the different technical systems and subsystems to combine for the different levels of control, needs to be acquired from different sources, for example from economical information systems within companies or general information from literature (Santos 1997).

A data acquirer meets difficulties, and needs to apply different methods when acquiring data on different parts of a product's life cycle. At level 1 it is generally possible to retrieve data from measurements, while, at levels 2 and 3, estimations and assumptions may have to be made, regarding local conditions of raw material extraction, suppliers, customers and waste managers, for example. This gives that the data on which decisions are to be based is of different quality, i.e. is of different reliability.

4. Difficulties with environmental information data quality

For each of the three different levels of control, the ability to identify environmental priorities accurately and efficiently, and the ability to navigate in the direction towards the most environmentally significant goals depends on the reliability of the environmental information. Also, the goals need to be identified and reached aptly, in order to avoid being faced with realised but unpredicted environmental damage or even catastrophes.

Therefore, an important issue for the environmental control function, at any level of control, is to efficiently acquire and maintain data quality, in means of its reliability. Reliability regards both the description of the technical system in subject, and numerical correctness of quantitative parameters of the system. Data quality maintenance is needed regardless of the type of data source, i.e. regardless of whether data is measured within a plant or whether it is modelled from different types of estimations and assumptions.

From the viewpoint of decision-making and control, the means for acquiring the data is irrelevant. The relevant issue rather is if the data, which is accessible to the decision-maker, sufficiently well corresponds with an ideally correct description of a technical system and an ideally correct numerical description of its quantitative parameters.

Regarding data quality maintenance, there is fundamental unfeasibility associated with assessment of quality of both models of technical systems and numerical correctness of quantitative parameters. The following sections will discuss aspects of this unfeasibility.

4.1 Unfeasibility of assessing correctness of models of technical systems

The description of any fairly complex real system is necessarily a gross simplification. During the modelling i.e. the simplification, some of the environmentally relevant properties regarding the system may not be included with the model. Excluded relevant properties may be in different states in the model and in the real technical system, causing the forecasting of the simplification to deviate from the actual outcome. The significance of the exclusions is difficult to appreciate in a decision-making process.

For more than fairly complex technical systems, such as product manufacturing, goods transportation or waste management, the difficulties with the modelling increase with also including difficulties with the system boundaries e.g. excluded or included subsystems or functionality may differ between the model and the real world representation.

Models of technical systems may be distinctly separated from any physical manifestation of the technical system they describe e.g. environmental reports may be stored in records at governmental authorities, or life cycle data, describing a production from cradle to gate, may be stored in a database containing general LCA data. Due to a decision-maker's ignorance regarding the technical systems and his lack of access to them for a correctness comparison, the reliability of the models may not be assessed. This has been studied in a project within CPM, where the feasibility to use environmental reports as a basis for reliable LCA-data were examined (Erixon/Ågren 1997).

4.2 Unfeasibility of assessing numerical correctness of data

To truly assess numerical correctness, data need to be related to a correct reference such as a calibrated numerical reference model. Such a numerical reference model may be designed from mathematical process models or mass balance calculations.

The unfeasibility with the approach of a numerical reference model lies in the difficulty with mathematically modelling a real technical system, with respect to realities, such as leakage, mechanical deterioration, mismanagement and other unpredictable influences to the quantitative environmental parameters of the system. It should be noted that this unfeasibility is also associated with the difficulty with finding a correct model of a technical system.

5 An approach to managing environmental information data quality - a data quality agreement

Due to the unfeasibility regarding assessment of correctness of both models of technical systems and of quantitative parameters of these technical systems, a different approach to data reliability is needed: both systems modelling and mathematical modelling strongly depends on subjective assumptions and choices of methods. This subjectivity suggests another approach for managing data reliability, instead based on organisational agreements and on structuring of routines.

5.1 The data quality agreement within SPINE@CPM

Within SPINE@CPM the agreement included a common view on how to interpret and handle environmental data, in terms of descriptions of technical systems (see figure 3). A set of documentation requirements, referred to as the *CPM data documentation criteria,* was commonly defined and agreed upon (Arvidsson 1997).

Basically, the agreement defines a description of a technical system to be considered reliable if it has been retrieved or attained using reliable methods. In other words, if the data is supplied with sufficient documentation describing these methods, the data user will be in a position to judge whether the methods used can be considered reliable or not, i.e. whether the data should be considered reliable or not.

There are several levels of methods to consider in regards of the design of a model of a technical system. A procedural approach for this design process has been developed at CPM. The model consists of six steps (Carlson/Pålsson 1998a):

0. Identification of conditions for acquisition of numerical entity
1. Acquisition of numerical entity
2. Statistical treatment of a set of numerical entities
3. Design of model of technical system
4. Aggregation of models of technical systems
5. Communication of environmental information on model of a technical system

The reliability of a method for acquiring numerical data depends on the methodological choices of each of the six steps: at step 0, the conditions for the acquisition, e.g. the state of a real technical system or modelling assumptions regarding this state; at step 1, the method for acquiring the numerical entity, e.g. the application of a measurement method or the modelling assumptions; at step 2, the statistical treatment of a set of numerical entities, e.g. the production of a frequency distribution or the identification of numerical boundaries of a subjective estimate and at step 3, the modelling resulting in the final design of a dataset describing a technical system, e.g. a detailed model based on causalities within a production system or a theoretical model of a technical system. Step 4 describes system analytical numerical treatment, such as LCA calculations or averaging. Step 5 describes data communication between different contextual environments, such as internal reporting and market communication.

Each step delivers a definite result, either simply numerical or as a complex description of a technical system. For each step the choice of method is arbitrary, within the range of available methods, while the resulting data is the result of the actual choice of method. The actual methodological choice needs to be sufficiently unambiguously documented and communicated to the following step, together with the resulting data.

It should be noted that the procedural description above is regardless of whether the methods are technical measurements or methods for estimations.

Note: Documentation of methods and other relevant information regarding data is in itself data, on a meta level, and are therefore generally referred to as metadata.

6. Means for fulfilling data quality agreement within SPINE@CPM

In order for the personnel at the CPM member companies to comply with the agreed data handling routines, they are first educated in the CPM data handling routines. Also they are supplied with a manual, specifically developed for the purpose and they are free to contact the CPM data administration at Chalmers University of Technology for questions.

Also continuos development of further methodological aids are being made within the CPM group. An example is the procedural approach for the design of a model of a technical system, described in section 5.1.

7. Maintaining data quality within SPINE@CPM

Since data reliability was considered dependant on documentation of choices and methods it was considered a risk that the system unnoticed could deviate from its original quality intentions. Therefore a review function, based on requirements on the documentation, was included with the system, with purpose to maintain the defined quality. The purpose of this function is to error-detect and to continuously reinforce conformance with the data quality agreement.

The review function is performed by a person well familiar with the requirements of the quality system. An important task for the review is to identify logical irregularities and decline in documentation quality.

It needs to be stressed that no „environmental knowledge„ of the reviewer is required, e.g. such as knowledge of amounts of emissions from different technical systems or special knowledge in ecology. This is due to the difficulties discussed in sections 4.1 and 4.2.

The review follows the procedure described in figure 5.

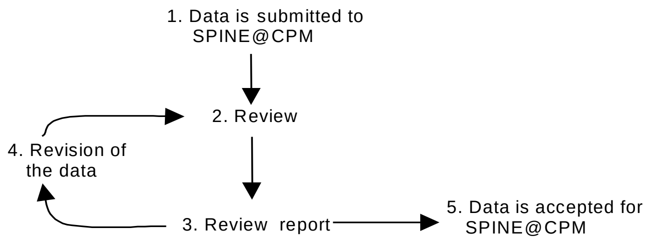

Figure 5

1. Personnel at the CPM member companies submit data to SPINE@CPM.
2. Review of the data is made in accordance with the CPM data documentation requirements, i.e. in regards of quality of the documentation of the model of the technical system and choice of methods.
3. A review report is written, which may contain notes and comments from the reviewer or requirements for supplemental information and corrections.
4. If the data is not accepted by the review, i.e. if supplemental information and corrections is required, the review report is sent back to the data supplier for further revision of the data. Steps 2-4 is repeated until the data is accepted by review.
5. If notes and comments of the reviewer require no additional revision, data is accepted for SPINE@CPM.

8. An environmental information system for data quality review

8.1 SPINE

The fundament of the SPINE@CPM information system is the SPINE data model and database structure. It is used to express the documentation criteria and the quality aspects of data (section 5), and it also defines the fundamental design of the software supporting this system.

SPINE was designed within a Nordic industrial project, holding a broad range of competence, including computing science, LCA and ecology. Participants came from about 15 large Nordic industry companies, Chalmers University and other competence groups. The result from the project was a technical report describing the model and which also includes a complete script for implementing the model into a relational database management system (Carlson et al 1995).

SPINE was designed as a relational database since it was planned to be the format for a national LCA database, but also because it was designed to bridge data-gaps between different types of LCA software and other software for environmental decision-support. It was intended that many different software should use the same database format for their data storage.

Today there are three fully developed and commercially available decision support software tools based on SPINE: EcoLab (Nordic Port), EPS Design System (Assess) and LCA iT (Chalmers industriteknik).

8.2 The technical functions of the SPINE@CPM information system

SPINE@CPM was developed within a project within the CPM group, in order to increase the accessibility of data on technical systems for LCA-research and industrial LCA-related applications. The aim of the project was to establish a national LCA database system. The project started in 1996 and it was finalised in the beginning of 1998 (Carlson/Pålsson 1998b). The system developed within the project is designed to support the data review function as described in section 7.

Figure 6 depicts the SPINE@CPM system technically: The database system is modularised, with the main module being the SPINE@CPM Data Tool, a software developed within the project. The data tool includes all functions of the system and all users within the system may use the same tool for their tasks.The tasks of the system are to

- enter data into the system
- quality review data
- move data sets between different SPINE-databases and different users.
- publish data at SPINE@CPM

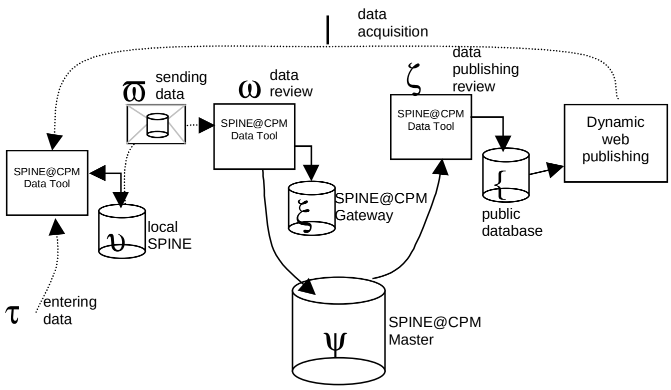

Figure 6

The data suppliers entering data (0), generally situated at the CPM companies, are equipped with a copy of the tool, together with a local database (1). The data review function (3) uses the tool for reviewing the documentation, and for moving accepted data sets between the received databases and the master database (5). Databases are transmitted from the data submitters via e-mail (2) to the review function at CPM. Review reports are also sent by e-mail, from the review function to the submitters (not shown in figure 6).

The system has one secure master database server (5), in which all quality-reviewed data are stored. The content of this database is never deleted, since the master database will also function as a reference, when comparing different copies of the same data set.

Data is published via Internet (7) using dynamically generated HTML-pages. These pages are generated from a local copy of the master database, both for security reasons and because it is not required to publish the entire content of the master database. Before any data is being moved from the master database to the public database the data is again being reviewed (6), in order to select which data to publish.

9. Conclusions and further development

In spite of the simplicity of the system, the difficulties with establishing it should not be underestimated. The difficulties with the harmonisation process far exceed the difficulties with both the reaching of the data quality agreements, and the development of the technical equipment. Examples of difficulties are secrecy, organisational motivation, rate of propagation of the understanding of the details of the agreement within the organisation and the establishment of the reviewer’s role.

In order for the system to reach a sustainable establishment within a corporation it needs to be integrated with the existing information systems and organisational routines. In addition to this, the organisational routine needs to be extended to also include responsibilities for the overall maintenance of the environmental information system. For example, an organisational position needs to be formulated regarding secrecy, extension of an understanding of the details of the agreement into the organisation and the role of the review.

10. Bibliography

Arvidsson, P. (ed.) (1997): Krav på datakvalitet CPM:s databas 1997, CPM-report 1:1997, Göteborg

Bresky, J. (1998): STORA Corporate Research, Personal communication

Carlson, R. et al (1995): SPINE A Relation Database Structure for Life Cycle Assessment, IVL-report B1227, Göteborg

Carlson, R. et al (1998): LCI data modelling and a database design, in: The International Journal of Life Cycle Assessment Vol. 3. No.2. pp. 106-113

Carlson, R., Pålsson, A-C. (1998a): A procedural approach to the design of a model of a technical system (under preparation)

Carlson, R., Pålsson, A-C. (1998b): Project report Establishment of CPM‘s LCA database, CPM report 1998:3, Göteborg

Erixon, M., Ågren, S. (1997): Miljörapporter som underlag till livscykelanalys, CPM-report 5:1997, Göteborg

Netherwood, A. (1996): Environmental management systems , in: Welford, R. (ed.) Corporate Environmental Management; Systems and Strategies, London

Pålsson, A-C. (1997): Handbok vid arbete med datakvalitet och SPINE, CPM-report 6:1997, Göteborg

Santos, R. (1997): Environmental Management Information Systems, Report 1997:11, Technical Environmental Planning, Chalmers University of Technology, Göteborg

Saur, K. (1997): Life Cycle Impact Assessment, in: The International Journal of Life Cycle Assessment, Vol. 2. No 2. pp. 66-70

de Smet, B., Stalman, M. (1996): LCI Data and Data Quality – Thoughts and Considerations, in: The International Journal of LCA, Vol 1, No. 2

Weidema, B., Wesnaes, M. (1996): Data quality management for life cycle inventories - an example of using data quality indicators, In: Journal of Cleaner Production vol. 4, no. 3-4, pp. 167-174