The ecoinvent database system: a comprehensive web-based LCA database

Rolf Frischknecht a,*, Gerald Rebitzer b

a Swiss Centre for Life Cycle Inventories, c/o EMPA, Überlandstraße, 129, CH-8600 Dübendorf, Switzerland
b Life Cycle Systems Group, Swiss Federal Institute of Technology Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Accepted 3 May 2005
Available online 22 June 2005

Abstract

The paper describes the general structure of the ecoinvent database developed by the Swiss Centre for Life Cycle Inventories. The database accommodates more than 2500 background processes often required in LCA case studies.

Quality guidelines, established in order to ensure coherent data acquisition and reporting across the various institutes involved, are described. These include aspects such as the reporting of pollutants (e.g., heavy metals), or the nomenclature of processes and elementary flows.

The data (exchange) format is also described. Processes are documented with the help of meta-information and flow data (including both unit process raw data and aggregated LCI results). The structure of the data format corresponds to the ISO/TS 14048 data documentation format. Data exchange between project partner institutes and between the database and its customers (database users) is based on XML-technology. Matrix inversion is used to calculate the cumulative LCA data using efficient algorithms and making use of the fact that LCA matrices are usually sparse.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Life cycle inventory (LCI); Ecoinvent; Data format; Matrix inversion; LCA database; EcoSpold; ISO/TS 14048

1. Introduction and background

Up to now, several public Life Cycle Assessment (LCA) databases exist in Switzerland, partly covering the same economic sectors [1–6]. However, life cycle inventory data for a particular material or process available from these databases often do not coincide and thus the outcome of an LCA is (also) dependent on the database used and therefore the executing institute. Furthermore, the efforts required to maintain and update comprehensive and high quality LCA-databases are beyond the capacity of any individual organization.

At the same time, LCA is receiving more and more attention in industry and authorities as one important tool for e.g., Integrated Product Policy (IPP), Technology Assessment or Design for Environment (DfE). In parallel with this increasing trend for LCA applications the demand for high quality, reliable, transparent and consistent LCA data increases as well. Only a few publicly available LCI databases fulfill these criteria, most of them from the 1990s and thus not up-to-date.

2. Goal of the ecoinvent 2000 project

Under the lead of EMPA, LCA-institutes in the ETH domain (Swiss Federal Institutes of Technology (ETH) Zürich and Lausanne, Paul Scherrer Institute (PSI) Villigen, Swiss Federal Institute for Environmental Science and Technology (EAWAG), and Swiss Federal Laboratories for Materials Testing and Research (EMPA) St. Gallen and Dübendorf) as well as the
LCA-department of the Agroscope FAL Reckenholz, the Swiss Federal Research Station for Agroecology and Agriculture, in Zürich started close cooperation. Together with the Swiss Agency for the Environment, Forests and Landscape (SAEFL or BUWAL), the Swiss Agency for Energy (BFE) and other Swiss Federal Offices, the Swiss Centre for Life Cycle Inventories was founded. The LCA-institutes launched the project ecoinvent 2000, which resulted in a harmonised, revised and updated Swiss national LCI database. The database comprises LCI data from the energy, transport, building materials, chemicals, paper and pulp, waste treatment and agricultural sectors reflecting the production and supply situation in the year 2000, based on the Swiss and European demand patterns. The major applicability is in a European context, but selected data sets (such as oil production in Nigeria or natural gas production in Russia) have a broader international application. The tasks within the project have been distributed according to the expert knowledge of the partners (see Table 1).

A large network-based database and efficient calculation routines are required for handling, storage, calculation and presentation of data and have been developed over the course of the project. The corresponding components are partly based on preceding work performed at ETH Zürich [7].

### 3. Basic structure of the ecoinvent database system

The ecoinvent database system consists of the following components (see Fig. 1):

<table>
<thead>
<tr>
<th>Database content</th>
<th>Responsible institute</th>
<th>Partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy supply</td>
<td>PSI</td>
<td>ESU-services</td>
</tr>
<tr>
<td>- Fuels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Heat production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Electricity production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials and waste</td>
<td>EMPA</td>
<td>Doka Life</td>
</tr>
<tr>
<td>- Plastics</td>
<td></td>
<td>Cycle Assessments,</td>
</tr>
<tr>
<td>- Paper and board</td>
<td></td>
<td>Chadacoff</td>
</tr>
<tr>
<td>- Basic chemicals</td>
<td></td>
<td>Öko-science</td>
</tr>
<tr>
<td>- Detergents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Waste treatment services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Metals</td>
<td>ETHZ UNS</td>
<td></td>
</tr>
<tr>
<td>- Wood</td>
<td>ETHZ ICB</td>
<td></td>
</tr>
<tr>
<td>- Building materials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Basic chemicals</td>
<td>FAL/FAT</td>
<td></td>
</tr>
<tr>
<td>Transport services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic chemicals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agricultural products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and processes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ad 1.** The central database contains Life Cycle Inventory data on energy systems, transport systems, waste treatment systems, chemicals, building materials, etc., and Life Cycle Impact Assessment (LCIA) methods such as the Swiss Ecological Scarcity 1997 [8], Eco-indicator 99 [9], IMPACT 2002+ [10] or the CML characterisation scheme 2001 [11]. The database is located on a computer server and accessible via the Internet.

**Ad 2.** Data are supplied by the partner institutes as non-terminated unit processes (i.e., they can and usually do contain exchanges from and into the technosphere as well as elementary flows). The computation of cumulative inventory results is performed with powerful calculation routines related to the central database. Unit process raw data as well as LCI results include (cumulative) uncertainty ranges.

**Ad 3.** The local database administrators of the participating institutes use the editor to create new data sets and to change, complete or delete existing data sets. The editor administrates the data set names (via a direct link to the central database, where the index of data set names is located), ensures the use of the actual list of names when compiling new inventories and includes a unit converter. The editor acts as the interface between the local administrator and the central database and generates files in the ecoinvent data format (named EcoSpold, see below).

![Fig. 1. The basic structure of ecoinvent database system.](image-url)
Ad 4. The administration tool supports the integration of data sets delivered by the cooperating institutes into the central database. It helps to verify the completeness of data sets, calculates inventories and (normalised and weighted, if appropriate) category indicator results and data sets, calculates inventories and (normalised and weighted, if appropriate) category indicator results and supports the administration of ecoinvent users.

Ad 5. The Query tool is the users’ interface to the database and is used to download data sets from the central database. It enables the search for individual processes, for processes of a certain economic sector (e.g., transport or energy sector) or for data from a certain institute. General information (so-called meta-information) on the processes (technology, age of data, geographic coverage, etc.) is accessible to everyone, whereas the quantitative LCI data is only accessible for registered ecoinvent members (customers).

Ad 6. The data exchange format lists all data fields that were available to describe a data set. It has evolved from the international SPOLD data exchange format [12] and corresponds to the international technical specification ISO/TS 14048 [13]. Some of the data fields are mandatory, i.e. information must be provided. Among other features, the data exchange format allows for specifying upper and lower estimates (or the 95% standard deviation) as well as the probability distribution (e.g., lognormal) of inventory data.

Ad 7. Commercially available LCA-software such as Emis, PEMS, Regis, SimaPro, TEAM and Umberto are used as local databases. These local databases are tailored for an implementation of ecoinvent data v1.0 and its updates. It is recommended to use the ecoinvent data (exchange) format for the purpose of data import.

In the following sections three aspects of the ecoinvent database system are highlighted. The first covers the quality guidelines for life cycle inventory analysis performed within ecoinvent 2000 project. The second describes certain aspects of the web interface (Query tool), and the third focuses on the calculation routines implemented in the ecoinvent database system.

4. Quality guidelines for ecoinvent data

The creation of one central life cycle assessment database requires a high degree of coordination and harmonisation. In this section, several harmonisation issues are listed and discussed. Beside structural aspects and naming conventions, content-related aspects have been elaborated and unified. This guarantees a maximum degree of consistency of the available process data.

4.1. Structural aspects and naming conventions

The inventory of a unit process is divided into the needed production means (infrastructure; including dismantling) and the operation phase. Thereby the entire production means necessary to run, for instance, a fuel cell combined heat and power unit is evenly spread over the expected life time production. For instance, 1 kWh “electricity, at fuel cell” requires a tiny share of 1 unit “fuel cell, 1.5 kW<sub>e</sub>”.

Co-production processes such as the above-mentioned fuel cell, which delivers both electricity and heat, are stored as such in the database (i.e., before allocating inputs and outputs to its co-products). In order to be able to attribute a certain share of requirements and burdens to each of the co-products, allocation factors are defined and stored separately. This procedure allows for a flexible and easy handling (and change/adaptation, if necessary) of allocation factors.

All processes are defined by the four data fields “name”, “unit”, “location” and whether it is an “infrastructure process” or not.¹ The process name is created according to the following order and scheme (see also Table 2):

1. Name of the good/service, production process or processed product, level of treatment (see below).
2. Additional description (if available) in the following fixed order: sum formula, site (or place of origin), company, imports included or not, any further description.
3. Level of the value chain (e.g., ‘at plant’, ‘at regional storehouse’) or destination (especially for wastes, e.g., ‘to waste incineration’, ‘to landfill’).

For an unequivocal identification of elementary flows, slightly different data fields are required, namely, name, unit, category and subcategory. The names of elementary flows are based on the structure developed by the SETAC working group on data quality and data availability [14,15]. Category and subcategory are used to describe the compartment (air, water, soil and resources) and its specification (e.g., for air: stratosphere, high population density, low population density, protected area and unspecified). Additionally, long-term emissions are distinguished for processes that are likely to emit during several thousands of years (such as landfill sites, nuclear waste depositories and overburdens at mining and milling sites).

Pollutants are reported only once and on the level of detail of the information source. This avoids double counting and conserves the original information and detail. For instance, benzene, reported as emitted to air in a highly populated area is registered under “benzene, air, high population density” but not under...

¹ For infrastructure processes, the construction (and dismantling) requirements and emissions are reported (as opposed to the other processes where only operational requirements (incl. a demand for a share of the infrastructure process) are reported).
“hydrocarbons, aromatic”, nor under “NMVOC, non-methane volatile organic compounds, unspecified origin”, nor under “VOC, volatile organic compounds, unspecified origin”. Benzene might, however, also be included in the sum parameter “NMVOC, non-methane volatile organic compounds, unspecified origin”, in case a Benzene emission is reported as NMVOC in the data sources used and the analyst is not able to identify the emitted hydrocarbon.

4.2. Content-related aspects

While structural aspects and naming rules can in many cases be controlled by the software system, a consistent implementation of content-related rules is less straightforward. Nevertheless, clear rules are required in order to minimise differences caused by individual, unsystematic choices of the LCI data collectors involved.

In ecoinvent, system boundaries are drawn based on expert knowledge and not based on fixed rules such as mass or energy shares. If an emission of a pollutant is expected but no data are available, estimates are used in order to identify whether or not this pollutant may be environmentally relevant.

Electricity is supplied as high, medium and low voltage with increasing losses and requirements of production means. Hence electricity demand of processes must be linked to the correct (or most likely) voltage. National supply mixes (as well as the export mixes) are calculated based on the domestic production plus the imports. In cases where the electricity mix actually purchased deviates from the average supply mix of a nation (or region) such specific mixes (or particular power plant technologies) are used.

In cases where the exact distances are unknown, standard transport distances are applied for the production phase of materials such as steel, cement, basic chemicals, etc. A similar approach is utilized for waste treatment processes. If no specific information is available, standard waste treatment processes defined per material are applied. It is assumed that inert materials go to landfills, plastics and paper are incinerated and metals are recycled.

Allocation is a ubiquitous issue, which calls for a harmonised approach. A cut-off approach is used for recycled materials and for by-products (outputs with no economic value that are not sent to waste treatment but are used in other processes). No burdens/impacts and no requirements of a preceding process chain and of a process are allocated to the recycled materials and by-products, respectively. On the other hand, no benefits are granted for any subsequent use of recycled material or by-product. No fixed prescriptions are made for joint product allocation (co-products) except that system expansion (especially the “avoided burden”-concept) is not recommended.

Fossil and renewable carbon are distinguished for CO₂-, CH₄- and CO-emissions. For renewable energy sources and materials, an equal amount of CO₂ is registered as a resource consumption according to the binding capacity of the corresponding crops. Carbon that is emitted as CO is considered when calculating CO₂-emissions. On the other hand, CO is assigned a global warming potential assuming its subsequent conversion to CO₂.

Uncertainty of flow data is quantified on the level of each individual input and output of the unit processes. If uncertainty is not known (because not stated in the sources used or because not known by the company providing the data) a standardised estimation procedure is used. A data quality matrix has been developed based on the pedigree matrix published by Pedersen Weidema and Wesnaes [16]. Scores from one to five are given for reliability, completeness, temporal correlation, geographical correlation, further technological correlation and sample size. Fixed uncertainty factors are attributed to each of the scores and an additional basic uncertainty is attributed to categories of exchanges (such as electricity and thermal energy consumption, groups of combustion emissions, waste treatment requirements and the like). In all cases a lognormal distribution is assumed. With the help of these standardised uncertainty factors, the geo-variance (square of the geometric standard deviation) is determined for each individual exchange in the unit processes.

5. Data exchange and XML-technology

5.1. Data set documentation

A process, its products and its life cycle inventory data are documented using the ecoinvent data format (EcoSpold) with the following structure:

```
Table 2
Examples of process names used in ecoinvent data v1.0

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Unit</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat, natural gas, at industrial furnace</td>
<td>RER</td>
<td>MJ</td>
<td>no</td>
</tr>
<tr>
<td>Gypsum, at plant</td>
<td>CH</td>
<td>kg</td>
<td>no</td>
</tr>
<tr>
<td>Electricity, low voltage, production</td>
<td>UCTE</td>
<td>kWh</td>
<td>no</td>
</tr>
<tr>
<td>Disposal, newspaper, 14.7% water, to municipal incineration</td>
<td>CH</td>
<td>kg</td>
<td>no</td>
</tr>
<tr>
<td>Transport, airplane, freight</td>
<td>GLO</td>
<td>tkm</td>
<td>no</td>
</tr>
<tr>
<td>Nuclear power plant, pressure water reactor</td>
<td>DE</td>
<td>unit</td>
<td>yes</td>
</tr>
</tbody>
</table>
```
Meta-information

Process

ReferenceFunction, defining the product or service output to which all emissions and requirements are referred.

TimePeriod, defining the temporal validity of the data set.

Geography, defining the geographical validity of the data set.

Technology, describing the technology(ies) of the process.

DataSetInformation, defining the kind of process or product system, and the version number of the data set.

Modelling and validation

Representativeness, defining the representativeness of the data used.

Sources, listing the literature and publications used in a data set.

Validations, listing the reviewers and their comments.

Administrative information

DataEntryBy, documenting the person in charge of implementing the data set in the database.

DataGeneratorAndPublication, documenting the originator and the published source of the data set.

Persons, listing complete addresses of all persons mentioned in a data set.

Flow data

Exchanges, quantifying all inputs from technical systems and nature to the process and all outputs from the process to nature and to other technical systems.

Allocations, describing and quantifying allocation procedures and factors, respectively, required for multi-function processes.

Customers (ecoinvent users) have the ability to check the content of processes online (see excerpt in Fig. 2). This helps to judge whether or not the process is of interest and worthwhile to download the corresponding data set. Links within the html-documents facilitate the navigation.

5.2. Role of XML-technology

Once a data set is chosen for download, one or several data sets are converted to one XML-file\(^2\) and saved on the local computer. XML schemes facilitate data exchange between different LCA-databases and -software. It can easily be extended by LCA-software-specific requirements and upwards and downwards compatibility poses no major problems.

For a flexible application of a data exchange between local LCA-software tools and the central database, a data exchange format in XML-technology is used. For that purpose XML schemes are applied. Although Document Type Definitions (DTD) are more widespread nowadays, they will most likely be substituted by schemes. On the one hand, schemes have a much higher performance and, on the other, they themselves use the XML language (as opposite to DTD, which uses its own specific language).

Schemes are used for validation and for documentation purposes. The scheme provides information on the general structure of an ecoinvent data set. Furthermore, all elements of a scheme may be completed with documentation information and comments.

6. Matrix-calculation with sparse matrices

A unit process is the smallest portion of a product system for which data are collected when performing a life cycle assessment [17]. Unit processes are linked by inputs and outputs with

- the technosphere (technical systems or the economy)
- the biosphere (nature or the life-supporting ecosystem).

The first link comprises exchanges with the economic system such as the purchase of intermediate goods (machinery, materials, ancillaries, etc.) or services (waste treatment, transportation, telecommunication). Direct resource extraction (e.g., supply of river water) and direct emissions to air, water or soil are relations of the second kind.

Accordingly, the directly caused energy and material flows of a unit process form a vector that is divided into a technosphere and a biosphere sector \(a\) and \(b\), respectively. The vector includes \(m\) technosphere exchanges \(a_i\) and \(n\) biosphere exchanges \(b_j\).

\[
\begin{pmatrix}
a_1 \\ \vdots \\ a_i \\ \vdots \\ a_m \\ b_1 \\ \vdots \\ b_j \\ \vdots \\ b_n
\end{pmatrix}
\]

(1)

The next step is to create a model of that part of the world economic system which comprises all processes required to produce a certain good or for the delivery of
a certain service. Such a model will — among others — include the energy supply sector, the transportation sector, the waste treatment sector and certain parts of the mining sector (minerals). All unit process vectors together form a matrix with an economic part $A$ and an ecological part $B$.

\[ P = \begin{pmatrix} A \\ B \end{pmatrix} \]

By calculating the inverse of the economic part $A$ of the matrix $P$ one gets the cumulative demand of intermediate goods and services required to produce each of the unit processes (economic part of the cumulative matrix called $C$). In order to be able to invert the matrix, it has to be square and non-singular. These two conditions are normally fulfilled when modelling real product systems.

The cumulative amounts of resource extractions and emissions (cumulative ecological part $D$) are calculated by multiplying the ecological part $B$ of matrix $P$ with the computed cumulative economic part $C$.

The inventory matrix $P^*$, composed of $C$ and $D$, contains information on the total (cumulative) requirements of economic entities (intermediate goods and services) and about the total (cumulative) flows of ecological entities (elementary flows, i.e. resource extractions and emissions) of all unit processes the matrix is composed of.

For the numerical implementation of the matrix inversion, direct methods are usually applied that make use of publicly available source code libraries (see [7] for further details).

7. Outlook

Since Fall 2003 consistent, quality-controlled LCI data with the reference year 2000 are available via the Internet for many basic commodities and services (such as energy supply, transportation and waste treatment services, building materials, wood products, chemicals and agricultural products) that are part of most LCI process networks. The database content will be updated and extended every two to three years. On which economic sectors to focus update and extension work is not yet decided.

The experiences gained concerning decentralised LCI data acquisition using common quality guidelines and its feeding into a central database are very valuable for future database developments. Variations of the data
exchange format are possible thanks to the flexibility of the XML-technology. Major LCA-software tool and LCA data suppliers are currently discussing the possibility of creating a common data exchange format that would further enhance data exchange and by that increase acceptance and efficiency of LCA. This may happen either within the life cycle inventory programme of the UNEP SETAC Life Cycle Initiative or within COST Action 530.

References


