

# Ecoinvent database version 3 – the practical implications of the choice of system model

**Bo Weidema**

The ecoinvent Centre, St. Gallen, Switzerland  
weidema@ecoinvent.org

**Abstract** In the new version 3 of the ecoinvent LCI database, the user can choose between different system models that reflect different LCI modelling algorithms applied to the same underlying unit process data. The system models reflect different consequential and attributional models, i.e. linking of inputs to either average or unconstrained suppliers, and arriving at single-product systems either by partitioning (allocation) according to different allocation properties or by substitution (system expansion). New approaches, such as the ILCD handbook's recommendation of system expansion with average flows, can also be accommodated by the new, flexible database structure. The presentation explains the new linking structure of the database that allows this flexibility and the structure of the workflow in database management. The practical implications for the end users of the different system models are highlighted.

## 1 System models in ecoinvent database version 3

Since its first publication in 2003, the ecoinvent database has become the most widely used background database for life cycle inventory (LCI) data. The harmonised guidelines for data collection and modelling developed by the project "ecoinvent 2000" have been an inspiration for many later national database guidance documents. After the successful update of the ecoinvent database to version 2 in 2007 it was decided to plan for a major re-structuring of the database for its version 3, due for publication late 2011, in order to ensure further expansion of its geographical and technological reach and detail, as well as enhancing its flexibility for use in different modelling contexts.

LCI modelling is composed of two elements: *Unit process modelling* and the linking of these unit processes in *system modelling*. The focus of the following is on the latter aspect, and especially the different LCI system models that will be supported by the ecoinvent database version 3 [1].

From the same unit process data, different LCI system models can be constructed that differ in two aspects:

- The linking of inputs to either average or unconstrained suppliers.
- The procedures to arrive at single-product systems in situations of joint production of products, which apply either partitioning (allocation) of the multi-product system into two or more single-product systems, or substitution (system expansion), which eliminates the by-products by including the counterbalancing changes in supply and demand on the affected markets.

A large number of different system models can be generated by combining these two modelling choices, which can be further modulated by varying the extent of constraints on suppliers and varying the allocation keys chosen for partitioning, respectively.

However, only a few combinations and variations are demanded in practice, where it is common to distinguish between consequential and attributional modelling:

- The *consequential* system model focuses on the long-term consequences of decisions, i.e. consequences that include changes in production capacity (capital equipment). This model therefore applies substitution (system expansion) consistently, and excludes suppliers that are constrained in their ability for capacity changes.
- The attributional system models all link inputs to the average suppliers, i.e. in proportion to their current production volumes, and therefore mainly differ in the way they arrive at single-product systems. In accordance with common practice, the ecoinvent database version 3 has one system model that consistently applies partitioning with the *revenue* obtained from the co-products as allocation key.
- A slight variation of this is a system model that applies instead "*true value*", which is a modification of the revenue to correct for situations where the revenue does not reflect the "true value" of the co-products. This system model is also planned to have an allocation correction for carbon, see below.
- Finally, the system model that the ILCD Handbook [2] recommends for decisions in the so-called *situation A*, where changes are not expected to affect production capacity, links the inputs to the average suppliers, except for suppliers of by-products, which are constrained. This system model also applies substitution, and is therefore only a slight variation of the consequential system model, but is nevertheless classified by its authors as an attributional model, probably due to its linking to average suppliers. However, unlike other attributional models the sum of the

product systems after substitution does not equal the system before substitution.

System models that apply revenue or true value partitioning have the drawback that they do not provide correct mass and elemental balances for the product systems. Systems that are partitioned will only be balanced for the property that is applied as allocation key. For this reason, it is planned to add an allocation correction for carbon to the system model with "true value" partitioning. An allocation correction is two datasets that counterbalance each other, re-allocating one or more environmental exchanges, so that the resulting allocated product systems have correct mass balances for the re-allocated exchanges. The rationale for applying the corrections only to carbon is that for carbon, in contrast to most other elements, the same substance as both input (capture of carbon dioxide from air) and output (carbon dioxide to air) has the same significant environmental impact pathway (change in the atmospheric concentration).

## **2 Database structure and modelling algorithms**

A central prerequisite for the new flexible database structure is the systematic introduction of separate names for activities and their products, and the introduction of market datasets for all products. It is through these markets that the inputs and outputs of all other datasets are automatically linked by the database modelling algorithms, unless the markets are deliberately circumvented by manually linking to specific suppliers. In its simplest form, a market dataset consists of a reference product, representing a consumption mix, and one or more inputs of the same product from the different transforming activities that are located within the geographical delimitation of the market. For each product at least one global market dataset is available. Depending on the product, this may then be sub-divided into geographically specific market datasets.

The algorithm for linking the intermediate inputs of an activity is to link to the local market activity dataset that supplies this input as its reference product. The local market activity dataset is identified by matching the geographical location of the activity with the available market for this location. Since markets do not overlap, there will generally be one and only one such market activity for each intermediate input. If the activity is defined for a geography that spans over more than one market, each of the market activities contribute in proportion to their production volume, implying that the intermediate input will be duplicated to match the number of supplying markets and the amount of the intermediate input will be divided over these in proportion to the production volume of each market.

The algorithm for linking the inputs of each market activity is to link to all those transforming activities within the geographical area of the market activity, which have the market reference product as an output, in proportion to their available production volume. It is here that different constraints on the availability of a supplying dataset can be taken into account.

The linking algorithms can specify one or more of the following constraints to be taken into account, leading to the exclusion of the constrained supplies from the market:

- *By-product constraints*, implying that only reference products (determining products) will be included in the supply.
- *Technological constraints*, implying that only suppliers with a specified technology level will be included as suppliers.
- *Market constraints*, implying that the corresponding input to the market is resulting from a reduction in consumption in another activity.

The consequential system model takes all three types of constraints into account while the ILCD situation A model only takes the first one into account. The other attributional models do not take any constraints into account.

The introduction of market activity datasets for all products implies that the naming of products become more important, because they define the markets and what products are included in the same consumption mix and thereby also what products can substitute each other. This implies that the same procedures that are relevant for defining the functional unit in a life cycle assessment is now relevant already when embedding an activity dataset into the database. Therefore, the description on how to correctly identify market boundaries and functional units (here: reference products) now has a much more prominent position in the ecoinvent Data Quality Guidelines [1] than hitherto.

For all system models, the linking algorithm identifies materials for treatment (by-product/wastes which are not provided as positive reference products of any other activity in the same geographical area), and move these to be negative inputs of the same activities, in order to include the treatment activities for the materials into the product systems. Since a negative input is the same as a positive output, this operation does not affect the mass, energy and monetary balances of the activities.

Finally, the linked, multi-product datasets are converted to single-product datasets, either by partitioning or through substitution.

Partitioning involves the generation of as many single-product datasets from each multi-product dataset as the dataset have products with the specified allocation property. For each of the single-product datasets, the original inputs and elementary outputs without the allocation property are multiplied by the ratio of the specified allocation property for the product (when multiplied by the amount

of the product) relative to the sum of this (multiplied) property for all outputs. This procedure is also known as co-product “allocation”.

Substitution is implemented by moving the by-products from being outputs of the multi-product activity to be negative inputs of this activity and linking this negative input to its local market, in the same way as described above for all other intermediate inputs. Note that by-products and wastes for which substitutes are not available have already been placed as materials for treatment by the procedure described above. This implies that for the remaining by-products there will always be an activity that supplies the by-product as its reference product, and which will therefore be displaced when an additional amount of the by-product from the multi-product activity is supplied to the market.

### **3 Workflow in the database management**

The new database structure implies a number of changes in the procedures for data providers and the database management.

Data providers no longer have to link datasets to suppliers, but only have to indicate the inputs by their product name.

In general, data providers no longer have to supply allocation factors, as these are now autogenerated from the relevant product properties, which in turn will be suggested by the relevant editors if omitted in a submitted dataset. The exception is if the price of a product is not reflecting its "true value", in which case the "true value relation" has to be specified as a property by the data supplier.

Data providers *do* have to specify one of the product outputs as the reference product (determining product), as opposed to a by-product/waste, which is normally straightforward. The ecoinvent Data Quality Guidelines [1] provide further guidance for situations where more than one product appears to be the reference product. As mentioned above, correct naming of intermediate outputs is now more important, since the name will determine which market the output will contribute to.

Data providers no longer have to specify whether an output is a by-product or a waste, since materials that require treatment are automatically identified by the database, as mentioned above.

Data providers are asked to consider specifying the technology level of transforming activity datasets as outdated, old, current, modern or new. If not specified by the data provider it will be assigned the default "current". Since the technology level is used by the linking algorithm to identify technology constraints, this may affect whether the activity is included or excluded from

contributing to the consumption mix of the consequential model. The specific algorithm for technology level depends on the market trend: When the production volume of the reference product is decreasing more than 3.33% annually, the activity is identified as unconstrained if its technology level is “old”, and when the production volume of the reference product is decreasing less than 3.33% annually, increasing, or stable, the activity is identified as unconstrained if its technology level is “modern”. If there are no supplying activities with the required technology setting, the requirement for “modern” is replaced by “new”, and “old” is replaced by “outdated”, and if these do not exist, by the option “current”. The 3.33% is derived from a generalised lifetime of 30 years for production equipment.

Data providers normally do not have to supply market datasets. If a new product is added to the database, a global market will be autogenerated by default. Only markets with more specific geographical locations or constrained markets have to be submitted separately. Constrained markets are modelled by adding a conditional exchange, i.e. an exchange that is only activated for the consequential system model, representing the amount of product that is resulting from reduction in consumption. The conditional exchange is added as a negative by-product output with the same name and unit as the reference product, and with a direct link to the consumption activity that supplies the reduction in consumption.

After approval by the review panel of the Editorial Board, the unlinked, multi-product datasets become part of the working version of the database, from which they are publicly available. By application of the above described linking algorithms, the unit processes are automatically linked to the rest of the database for each of the system models. The linked unit processes and the fully linked systems are then ready to be provided to LCA software users, without any requirements that the software have implemented the linking algorithms. Finally, matrix inversion is applied to the linked systems so that the accumulated LCI results for each system model can also be provided to end users without own LCA software.

## **4 Practical implications for end users**

With the new flexible database structure the conscious choice of system model comes within reach of the ordinary database end user. It also becomes possible to compare results of modelling the same product in different system models, with the same underlying data. This means that discussions on choice of system model can be separated from discussions of differences in data between the models.

It should be noted that the choice of system model is not an arbitrary choice. Different system models may be relevant for different purposes, while for a specific purpose typically only one model is relevant. Thus, not all system models may be of interest to all users.

Different system models can vary significantly in their results for the same products. The data for average suppliers may be very different from the data for a modern or old supplier. For example, chlorine production by the mercury cell process (old), the diaphragm cell process (current), or the membrane cell process (modern) gives significant different results. As activity datasets describing average production conditions are increasingly being replaced by separate datasets for current and modern technologies, the difference between models using average suppliers and models using unconstrained suppliers can be expected to increase. Also excluding a by-product from an average, as in the ILCD situation A model, can be significant when the by-product makes up a large part of the market, as for example in the markets basic metals like aluminium and steel.

## 5 References

- [1] Weidema B P, Bauer C, Hischer R, Mutel C, Nemecek T, Vadenbo C O, Wernet G, *Overview and methodology. Data quality guideline for the ecoinvent database version 3*, Ecoinvent Report 1(v3), St. Gallen: The ecoinvent Centre, 2011
- [2] European Commission, ILCD Handbook. General guide for life cycle assessment - Detailed guidance, Ispra: European Commission Joint Research Centre, Institute for Environment and Sustainability, 2010.