Environmental Modelling & Software 86 (2016) 204–218

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

# Contribution-based prioritization of LCI database improvements: Method design, demonstration, and evaluation

Juergen Reinhard [a, e, *], Christopher L. Mutel [b], Gregor Wernet [c], Rainer Zah [d], Lorenz M. Hilty [a, e]

[a] University of Zurich, Department of Informatics, Informatics and Sustainability Research Group, Binzmuehlestrasse 14, 8050 Zurich, Switzerland
[b] Laboratory for Energy Systems Analysis, Paul Scherrer Institute, PSI, 5232 Villingen, Switzerland
[c] Swiss Centre for Life Cycle Inventories, Ecoinvent, Lerchenfeldstrasse 5, 9014 St. Gallen, Switzerland
[d] Quantis Serl, Reitergasse 11, 8600 Zurich, Switzerland
[e] Empa Materials Science and Technology, Technology and Society Laboratory, Ueberlandstrasse 129, 8600 Duebendorf, Switzerland

ARTICLE INFO

*Article history:*
Received 1 April 2016
Received in revised form
15 July 2016
Accepted 28 September 2016

*Keywords:*
Prioritization
Life Cycle Inventory database
Meta analysis
Contribution analysis
Life Cycle Assessment

ABSTRACT

The largest Life Cycle Inventory (LCI) databases contain about 10,000 unit process datasets. Assuming that updating one dataset would, on average, require one working day, then updating the entire database would roughly require the continuous work of five workers for ten years. Methods are therefore needed to prioritize datasets to be updated, and such a method should be able to identify the unit processes that contribute the most across different Life Cycle Impact Assessment (LCIA) methods. To date, such prioritization methods are not available.

This paper presents an operational prioritization method. We demonstrate and evaluate our method by applying it to the ecoinvent database. The case study shows that prioritization of improvement efforts is highly useful because a robust nucleus of unit processes proves to be consistently important across all product systems and many LCIA methods. Focusing research efforts on these processes allows the effective improvement of the LCI database.



## 1. Introduction

### 1.1. LCA: background data is key but uncertain

In our modern economy, the provision, use, and disposal of almost any product involves complex globalized networks—called product systems—consisting of thousands of interlinked human activities (Hellweg and Milà i Canals, 2014). Accounting for the cumulated environmental impact of such product systems requires tracing and measuring the metabolism of all activities associated with a particular product. Life Cycle Assessment (LCA) is one of the most prominent techniques developed for this purpose (ISO, 2006). It focuses on the “compilation and environmental evaluation of all processes, outputs and the potential environmental impacts associated with a product system” (ISO, 2006, p. 2) with the goal of pinpointing ecological weaknesses, comparing alternatives, evaluating the main environmental impacts, designing new products, measuring the environmental relevance of a material or product, and establishing recommendations for actions (Guinée et al., 2001).

The LCA procedure is structured around a static model viewing the product system as a network of basic building blocks called unit processes. A unit process represents one specific activity or a group of activities in the product system and records (i) the *intermediate exchanges* from and to the technosphere, i.e., the input of usable energy and raw materials and the output of products and waste and (ii) the *exchanges with environment*, i.e., the input of natural resources and output of emissions (Finnveden et al., 2009). The unit processes are linearly linked by means of their intermediate exchanges. The calculation of the product system for the reference flow of interest results in the sum of all *exchanges with the environment* for the entire life cycle, called Life Cycle Inventory (LCI). Life Cycle Impact Assessment (LCIA)—the next

* Corresponding author. University of Zurich, Department of Informatics, Informatics and Sustainability Research Group, Binzmuehlestrasse 14, 8050 Zurich, Switzerland.

*E-mail address:* juergen.reinhard@empa.ch (J. Reinhard).

http://dx.doi.org/10.1016/j.envsoft.2016.09.018
1364-8152/

step within the LCA procedure—then takes the inventory data on these exchanges as an input to determine the impacts on the environment.

Although this procedure reduces the complexity of the real-world system under study, LCA remains a very data-intensive methodology. A typical product life cycle covers thousands of unit processes, each of which needs to be described with exchange flow data. This information can usually not be gathered within a specific project due to the high cost of primary data collection. It is therefore common practice to focus data collection efforts on selected activities that reflect the space for action—these activities are together called the foreground system—and to use generic data from Life Cycle Inventory (LCI) databases to model the remaining activities, called the background system (Bourgault et al., 2012; Tillman, 2000). The background system usually covers up to 99% of the unit processes in the product system; only in rare cases does the number of unit processes modeled explicitly in the foreground system exceed 5%. Bearing this in mind, background or Life Cycle Inventory (LCI) databases can be considered the backbone of any LCA study. They provide the dominant share of the building blocks required for any LCA: aggregated and/or disaggregated unit process data. Therefore, the available quantity and quality of unit process data provided by LCI databases are of utmost importance.

However, the unit process data provided by LCI databases are characterized by a high degree of uncertainty (Groen et al., 2014). This has several reasons. First, it is not feasible to match each original activity in the real world with a description of a unit process in the LCI database. Consequently, the exchange flow values recorded in a unit process often represent assumed average conditions of a whole country or larger region across a given time period and across different instances of real processes. However, industrial production characteristics and environmental conditions can vary significantly over geographic space, time, and process instances (Huijbregts, 1998). Second, the quantitative data required to compile an accurate representation of a specific activity may be unavailable, wrong, or unreliable (Ciroth et al., 2004; Heijungs and Huijbregts, 2004). Third, the growth of LCI databases during the last twenty years was primarily supported by a local funding structure. Consequently, LCI databases are biased in their coverage of real-world activities towards their local origin, whereas supply chains are global. Although current developments improve on this shortcoming (see for example Wernet (2012)), today's LCI databases are far from being complete, even on the level of generic data.

Heijungs and Huijbregts (2004) present various approaches to address these uncertainties;

- the statistical approach (applying methods from statistics to existing LCI data),
- the constructivist approach (involving stakeholders),
- the legal approach (“truth” decreed by authoritative bodies), and
- the scientific approach (doing more research to create or improve primary LCI data).

While all approaches have merit, this paper explores a combination of the statistical approach with the scientific approach: Statistical methods provide the focus for the scientific approach in the systematic improvement of quality in LCI databases.

### 1.2. Background databases: effective prioritization is challenging

We believe that a statistical approach has particular merit because effectively prioritizing improvement efforts is becoming increasingly difficult due to the growing numbers of unit processes stored in existing databases. For example, depending on the system model,[1] the current version of ecoinvent (3.1) includes between 10,305 and 11,332 unit processes. If updating one unit process would, on average, require one worker day,[2] a systematic update of the entire database within one year would require the continuous work of more than 50 people. Capacities for such extensive improvement efforts are typically not available. In addition, the desired quality and quantity of databases is a “constantly moving target.” It is the “continuing evolution in consumer preferences, market and industry imperatives, and public policy which forces continuous development and improvement of datasets and methodologies for LCA to meet these needs” (Sonnemann and Vigon, 2011, p. 98). Compliance with these developments typically ties up a lot of the limited workforce. Consequently, improvements in data quality and quantity are so far rather directed by external data availability and situation-driven requirements than by systematic choice. To our knowledge, none of the LCI databases applies statistical methods to prioritize LCI database improvements.

Moreover, LCI databases represent highly interconnected systems, where almost every unit process is involved in every product system. As noted by Heijungs (2012, p. 172), “due to the interlinked nature of many processes, seemingly alien[3] parts of the database are involved in all LCAs”. This makes it difficult to align improvement efforts effectively to the unit processes that matter most or to anticipate the changes induced by an update in advance. For example, each of the roughly 4000 product systems in ecoinvent v.2.2 require between 2100 and 2300 processes (Heijungs, 2012). That is, the system-wide representation of almost every product system involves more than half of the unit processes in the entire database (Heijungs, 2012). Vice versa, updating an existing or integrating a new unit process can affect the results for a large part of the products. Identifying the unit processes that matter most would allow untangling these complex interdependencies and facilitating the effective and efficient improvement of the entire LCI database.

### 1.3. Prioritization of LCI database improvements

In general, we can distinguish two perspectives to prioritize LCI database improvements: external and internal prioritization. External prioritization aligns inventory efforts to importance based on an external reference such as environmentally extended input-output (EEIO) databases. For example, Majeau-Bettez et al. (2011) contrasted the economic, environmental, and structural importance of the economic sectors (obtained from an EEIO database)

[1] A system model provides a set of rules that specify how activity datasets are linked to form product systems. Ecoinvent version 3.1 offers three system models to choose from. They differ mainly with regard to their handling of the multi-functionality problem (system expansion vs. allocation), the use of average or marginal suppliers, and their assessment of by-product treatments (Weidema et al., 2013).

[2] The workload for updating datasets varies greatly. Depending on the given level of completeness of a dataset, it might take weeks (e.g. when a global dataset is disaggregated into many country specific datasets) or just an hour (e.g., when the quantity of one emission flow in one dataset is updated). Updating a dataset typically involves data collection, data entry/manipulation and data submission to peer review (using a software tool called EcoEditor). Any change in existing datasets or the submission of new datasets requires a peer review to be accepted into the database (Weidema et al., 2013). Therefore, one worker day should be considered as a rough but realistic estimate of the average effort associated with updating a dataset.

[3] For example, the product system associated with the provision of 1 kWh of high voltage electricity (produced by a hydro power plant in an alpine Swiss region) involves “alien” processes such as airport infrastructure, sugar (from sugar cane), kraft paper, and many more.