SCORE LCA

STUDY N° 2014-06

HOW TO COMPLETE LCA STUDIES DATA COLLECTION IN ORDER TO IMPROVE ITS QUALITY?

English scientific summary

November 2015

Scientific manager :

– Etienne LEES-PERASSO, Julie ORGELET, Axel ROY

Bureau Veritas CODDE

170 rue de Chatagnon – 38430 MOIRANS

SCORE LCA is an association that has been created to financially support collaborative research on LCA and related topics. It aims to promote and organize cooperation between companies, institutional and scientists in order to support the evolution of LCA methods and its practical implementation at European and international level.

- ✓ This work has been supported by ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie) www.ademe.fr

- ✓ The views and recommendations expressed in this publication are those of the authors and do not necessarily reflect, unless otherwise stated, the views of all members of SCORE LCA.

- ✓ The information and conclusions presented in this document were established on the basis of scientific and technical data and regulatory and normative framework in force at the date of the publication of documents.

1. Summary

LCA is a more and more common and widespread tool used to calculate the environmental impacts of products and services. In addition to an analysis tool, it is now a decision making asset: technical solutions choice, supplier selection or even regulatory decisions, LCA has a growing importance. **This importance leads to an increasing need for reliability and transparency of data and results.**

In parallel, the multiplication of LCA studies towards different goals (environmental labelling, ecodesign, etc.) as well as the growing complexity of reference documents and analyzed systems leads to **an increasing need of temporal and human resources.**

Most resources allocated to LCA studies go towards the collection and research of data to assess the systems considered:

- **Specific (or primary) data**
- **Generic (or secondary), or semi-specific data**

The time needed to perform this collection and the assessment is even greater as some data may be missing, hard to find or even false. Therefore, it is required to define methodological choices in order to fill those gaps. Those choices are essential and must preserve the relevance of studies. Indeed, in some cases, the environmental impacts related to missing data represent a non-negligible part of the overall impact of the system studied.

Therefore it is necessary to have methods to fill the missing data gap available, allowing altogether to limit the need for resources as well as ensuring the results reliability and transparency.

This study aims at answering this need through three steps:

- Description of existing reference documents exigencies in terms of missing data
- Identification, description and application of actual methods or methods in development aiming at solving the data missing gap
- Determination of recommendations allowing the selection and application of chosen methods, depending on the studied system typology and the LCA study goals.

It leads to both the conclusions about the use and evolution of exigencies and methods, as well as recommendations and tools related to the selection and application of methods while controlling the associated risks.

The main conclusions of this study are:

Concerning the exigencies, the analysis enlightened the different logical schemes (or the lack thereof) concerning their implementation in the different documents, depending on:

- **The activity sector:** adaptation of some exigencies to specific sectors in order to reflect specific constraints
- **The age of documents:** there is no apparent logic, but this point could evolve shortly due to the actual harmonization around two main reference documents: the ILCD handbook on one side and the EN 15804 on the other.
- **Involved actors and documents objectives:** there is an opposition between two main objectives: LCA applicability and democratization on one side, using low-constraint exigencies in terms of resources, and the methodological relevance on the other.

  Another aspect concerns the importance given to exigencies for databases. They often have many and strong exigencies. It mainly allows their users to use their data for many different applications using different reference documents, so they have to comply with all those exigencies.

This study shows that the actual developments go towards a greater harmonization and complexity of exigencies.

Concerning the methods, many exist, covering quite well the need for identification, analysis and substitution of missing data. Though, those methods have different application perimeters and advantages and drawbacks, due to inherent or external reasons. The recommendations and tools developed in this study are therefore focused on this aspect.

Methodological evolutions (reference documents, standards and regulations) are actually in progress concerning the usage of methods, but there are only few works concerning the elaboration of new methods.

On the other hand, each sector and each firm have their own stakes and problematics, hence a different usage of reference documents, exigencies and methods. There is then a need of appropriation and capitalization on those aspects.

To go further, it is important to gather and analyze feedbacks from the different stakeholders (practitioners, verifiers, study users), as well as the methods and exigencies answers to studies objectives. This could provide a clear synthesis to be used as a basis for the development or update of future reference documents.

2. Work and results presentation

1. Phase 1: Standards and reference document exigencies related to missing data

The phase 1 of the project firstly concerns the exigencies identification and description, then their individual and comparative analysis in order to determine their ease of use and relevance, as well as the induced need of resources from practitioners.

To perform this task, all identified exigencies are analyzed in a table to provide a global overview, and to allow merging based on different criteria, which leads to a first analysis approach.

In total, this identification leads to the analysis of **23 documents**:

- 9 databases
- 6 general and sectorial standards
- 5 environmental declaration programs
- 3 reference documents

For each document, all exigencies related to missing data has been noted and regrouped in different categories based on their similarities.

In total, **72 exigencies** have been identified, regrouped in **6 different categories**:

- 11 checklists
- 14 documentations
- 6 missing data identification
- 2 iterations
- 17 cut-off rules
- 22 substitutions with generic / average / penalizing data

Those categories are as follows:

- **Checklists**
    - Definition: elaboration of a list of steps to check during the collection or verification phase, in order to ensure the completeness of information on some critical aspects.

  Many different checklists exist depending on the reference document and product typologies. They can be defined in the documents themselves (as for some environmental declaration programs) or before the data collection, thanks to the capitalized knowledge on similar systems.
- **Documentation**
    - Definition: aims at filling out the methodological choices and identified missing data in a transparent way.

  This category includes very diverse exigencies: some of them only define that it is necessary to indicate which data are missing, others require specific communication means to do so.
- **Missing data identification**
    - Definition: aims at determining whether the studied system is complete, or if there are some missing data. It also aims at identifying the origin and/or the nature of those data.

  Several kinds of exigencies exist:

  - The creation of a flow chart before the data collection
  - Mass, energy or other physical or virtual data balance checks
  - Results plausibility checks

- **Iterations**
  - Definition: consists in performing the data collection again in continuous improvement logic, aiming at completing missing data that were not identified or which could not be filled out before.

- **Cut-off rules**
  - Definition: consists in the exclusion of a certain percentage of the system from the assessment.

  The cut-off exigency principle is very common, but in its application many particularities exist depending on the reference documents:

  - Several cut-off levels (5%, 1%, 0%)
  - Several criteria (mass, energy, environmental impacts, etc.)
  - Several ways to check the compliance: sensitivity analysis, substitution with generic data, etc.

- **Substitutions with generic / average / penalizing data**
  - Definition: replacement of missing data with other, generally non-specific, data. They could be from different origin and nature depending on the exact exigencies and study objectives.

  Three kinds of substitutions are employed differently by each exigency:

  - Generic data: use of non-specific data or values from databases, or from calculus or estimations
  - Average data: use of local or international averages
  - Penalizing data: use of detrimental hypotheses, mainly in order to ensure the real impacts are lower.

For each of those categories, the identified exigencies have their specificities, advantages and drawbacks.

Finally, a global analysis has been performed in order to enlighten the potential logical schemes underlying the exigencies choice and the different documents.

Several analysis levels have been considered, based on the exigencies Excel file. The conclusions are as follows:

- **Activity sector:** no apparent logic scheme. Indeed, several sectors have a good maturity (as seen in the previous step 2.1 about the identification and characterization of missing data); hence they have similar problematics concerning the missing data.

  Though, some exigencies, such as the checklists, are adapted to each sector in order to reflect the specific constraints. Moreover, some exigencies are more or less applicable depending on the activity sector. A good example is the migration of the building sector exigencies to the electric and electronic sector with the EN 15804 standard.

- **Age of the documents:** no apparent logic scheme. The strongest exigencies do not necessarily come from the most recent documents. This is mainly due to the fact that for a long time most documents have been based only on the main ISO standards (1404x and 1402x series), plus some sectorial specific documents. Though, this could change shortly because of the actual harmonization tendency towards two main reference documents: the ILCD handbook on one side (being used as a reference by the PEF scheme) and the EN 15804 on the other side (being used by the DHUP decree, the XP C 08-100 standards and some environmental declaration programs).

- **Involved actors and documents objectives:** there is a logic scheme, mainly visible through the opposition of two main objectives: on one hand the democratization and accessibility of LCA, using low-constraint exigencies in terms of resources, and on the other hand the methodological reliability. This difference is particularly visible through the analysis of two documents: the PEF on one side and the UNEP work on the other. Their conclusions diverge, mainly on the cut-off exigency.

  Another aspect concerns the importance related to those exigencies for the databases. Indeed, they have a double constraint: first they must allow their users to use their data for a wide range of applications; hence they must respect the exigencies of many reference documents. On another side, the large number of existing data leads to creation and maintenance difficulties if the exigencies are too strong.

To conclude, it is necessary to replace the exigencies analysis in the current context (growth from a national to a European or worldwide scale), and through the different actors objectives (databases, regulatory needs). Globally, the ongoing works move towards a greater harmonization and more complex exigencies.
To those verification and missing data completion exigencies, there are additional requirement related to the documentation and documentation format, mainly the ISO 14048 and the ILCD handbook.

2. Phase 2: Existing and developing methods summary

This second phase lists and analyzes actual and ongoing methods aiming at identifying, limiting or substituting missing data.

Step 2.a: Identification and characterization of missing data

This step aims at defining what missing data are, and from that where are the difficulties met by the LCA practitioners. It allows the classification of the different data types and to check their connections to the different activity sectors. This analysis leads to a mapping of difficulties met: what are the most missing data, and where are they?

First, it is important to define what missing data are.

There are two types of data:

- Product of waste flow data type
- Elementary flow data type

They also have two attributes:

- The quantitative information: quantity consumed or emitted
- The qualitative information: nature of the flow, the material, the process…

In this study, data are called missing if one or several of those types or attributes are absent or incomplete.

Once the missing data notion defined, **we seek to identify in which sectors the difficulties are.**

To do that, the data attributes have been cross checked with the different activity sectors concerned by the LCA realization. We chose to use **the NACE classification** as it is recognized and exhaustive. It led to the identification of the sector/attribute combination for which the data access is the most difficult, or on the contrary the most mature sectors.

The global analysis shows **there is a correlation between the quantitative and qualitative information coverage**:

- When the qualitative information is entirely or partially covered, the quantitative information is present, or at least partially present. It can be explained by the fact that life cycle inventory data sets are generally based on the available quantitative information. If some information is present in the sector, it facilitates databases establishment.

- On the contrary, when the qualitative information is missing, it leads to a missing, or partial quantitative information. Here as well it can be explained by the necessity to use qualitative information during database creations

Yet, this correlation only works in one way, from the quantitative to the qualitative data. Indeed, the qualitative data presence doesn't necessarily imply the quantitative data presence. The creation of statistical data is not necessarily linked to the life cycle databases creation, and is generally related to other uses, such as regulation, normative or commercial.

This analysis shows that the presence of quantitative data is a prerequisite to the determination of life cycle inventory database used for life cycle assessment studies. This could direct the new LCI data sets development towards new activity sectors, favoring the establishment of preliminary statistical studies that could reveal necessary.

Step 2.b: Methods definition

This second step aims at establishing a list and description of existing methods and methods in development, and to analyze them concerning their relevance and advantages and drawbacks.

Methods have first been identified from the exigencies listed in phase 1. Indeed, many exigencies are referring to the application of a specific method (material balance, data substitution, etc.). In addition, a documentary analysis has been performed to identify complementary methods, actual of in development.

In a similar way to the exigencies analysis, all identified methods are listed and detailed in a table.

In total, 20 methods have been identified, regrouped in **7 different categories**:

- 3 results relevance and coherence verification
- 5 input / output balance
- 3 checklists
- 1 missing data quantity estimation
- 3 data collection planning
- 1 reverse engineering
- 4 substitution with estimated / generic / average / penalizing data

Those categories are as follow:

- **Results relevance and coherence verification**

  The objective is to check the relevance and coherence of the missing data choice, in order to validate the data collection work, related to the objectives and perimeter of the study.

- **Input / output balance**

  It is about considering the system (or a part of the system) as a black box for which inputs and outputs must be equivalent.

  Those balances are based on mass, energy or element conservation principles, or other physical or monetary flows.

- **Checklists**

  It consists in a list of tasks, phases, steps, established in order not to forget any important element.

- **Missing data quantity estimation**

  This method proposes missing data quantity estimation, as well as the corresponding uncertainties. It consists in assessing the processes with statistical estimators with properties adapted to small size samples and highly variable data.

- **Data collection planning**

  The development and planning of the data collection consists in the definition of roles and responsibilities of each actor, as well as the perimeter and the collection process,

the collected data documentation and management, the collection effort and the assessment granularity for the different elements of the study.

This planning can and must evolve as the data collection and processes are refined.

- **Reverse engineering**

  This method consists in the determination of the manufacturing process of a system, as well as in identifying some elements and materials, from technical information available from the product or the observation of the production line. It is then possible to use that information and confront it to the data collection in order to identify missing data, or to define the data collection file, upstream.

- **Substitution with estimated / generic / average / penalizing data**

  In order to substitute missing data, several data types can be used, depending on their availability and the study objectives: estimated, generic, average or penalizing data. Each one has specific applications and consequences.

Step 2.c: Determination of the methods limits on actual case studies

This step focuses on the practical application of those methods. The work performed in this analysis allows the completion of the theoretical approach of the previous step with a practical one. In detail, the studied aspects are:

- Effect on the impact results and / or the study conclusions
- Easiness of application
- Resource needs

To do that, each method has been tested on actual case studies.

Three types of case studies have been assessed:

- Mono-material or simple multi-material products: product constituted of few materials and with a simple production process.

  This type mainly allows the determination of the influence of methods on the environmental impacts thanks to a simple, well-known and understood case.

  Used data: LCI data of 1 kg of galvanized steel production – *Worldsteel* – 2011

- Complex manufacturing products: products with few materials, but with a complex manufacturing process and / or supply chain.

  This type allows the evaluation of potential phenomenon of the methods impact propagation through different phases of an intertwined system.

  Used data: French electricity production mix data (Market for electricity, low voltage) – *ecoinvent*, allocation, ecoinvent default, v3.1 – 2008

- Complex products, assemblies, systems: those products are mainly characterized by a large quantity of input materials and components, causing specific data collection for each of them difficult because of the high number of data.

  This type allows the confrontation to resources need in terms of time and human resources, as well as to the methods application limits.

  Used data: 8,067m² Spanish low-energy residential building construction case study - *Use of LCA as a Tool for Building Ecodesign. A Case Study of a Low Energy Building in Spain* – I. Zabalza, S. Scarpellini, A. Aranda, E. Llera, A. Jáñez – 2013

This overall analysis, theoretical as well as practical, lead to the determination of advantages and drawbacks associated with each method, and on a more global scale with each method category.

Detailed information for each method is also indicated in the method sheets given to ScoreLCA members (see Phase 3: Practical and concrete recommendations).

This work showed that there are two sides to the methods: an inherent side concerning their ease of use, advantages and drawbacks and associated risks, but also an external influence, dependent of studies constraints: objectives, used reference documents, application time, etc.

This synthesis details the second side.

The different analysis levels are:

- **Application time phase:** we differentiated 3 consecutive phases of LCA for which the application of methods related to missing data is relevant:
    - **Upstream:** this preparation phase corresponds to the definition of the perimeter and objectives of the study, and also to the choice of databases and software used. During this phase, it is important to apply methods concerning the planning and anticipation of the data collection needs.
    - **During the LCA:** this phase corresponds to the inventory (data collection), assessment of the systems and interpretation steps.
    - **Downstream:** this phase corresponds to the verification (internal or external) during which the methods application influence is checked, in addition to the estimation of the quantity and importance of missing data.
- **Study type**
    - **LCA study:** those studies concern the environmental analyses, during which some methods can be applied. As for the UPR, the assessment details are known, but the data collection and assessment can have a more variable quality, depending on the objectives and perimeter of the studies.
    - **UPR Unit Process:** “smallest element considered in the life cycle inventory analysis for which input and output data are quantified.” *[ISO 14040]* This format is generally the one precising the most information on the assessment and hypotheses, because of the need for transparency when using those data. Its construction allows going back to the model in order to understand and adapt the data depending on the user needs.
    - **LCI:** By this term, we consider the life cycle inventories aggregated at the system input / output flow level. This format is common in databases, such as ELCD. Its advantages are its ease of use, and the fact it can hide confidential information, but the user loses some information on the assessment.
- **Objective type:** for this analysis level, most methods can be used indifferently whatever the objectives are. Yet, there are specificities for some methods:
    - **Environmental declaration / communication:** in this category are included LCA realized in order to communicate the impact results or the environmental benefits associated with the studies system to a third party, external to the firm (client, general public, purchaser, etc.). The comparative LCA destined to the general public are excluded of this category.
    - **Ecodesign:** in this category are included studies performed in order to analyze or improve the system within a company or a company network, without external communication.
    - **Comparative assessment:** in this category are included studies performed in order to position a system within other system, internal or external to the company, on the environmental level. The comparison can have communication aims.

To conclude, many methods actually exist and cover the missing data identification, analysis and substitution needs quite well. Yet, those methods have different perimeters and applications, as well as advantages and drawbacks from inherent or external reasons, which must be apprehended before performing an LCA.

The phase 3 deliverables go from the conclusions of this analysis and summarize them in order to give LCA practitioners tools and procedure sheets to select and apply methods appropriately depending on the criteria of each study.

3. Phase 3: Practical and concrete recommendations

Following the analysis work on exigencies and methods, this study dealt with the elaboration of recommendation on the selection and use of methods depending on the LCA practitioners' different constraints. Indeed, the main difficulty lies in the choice and application of the methods: which one to use at which time to cover my need, and what are the associated risks?

The objective is then to provide simple tools making this selection and use easier, and *in fine* to improve the LCA studies quality.

We answered in French this problematic in three steps:

- **Elaboration of a procedure sheet** summarizing the main questions to deal with upstream, during and downstream an LCA: « Score LCA_Procédure_20151007_final »
- **Elaboration of method choice matrices** defined in function of determining criteria, and enabling the practitioners to select the relevant methods: « ScoreLCA_Matrice des méthodes_20151102_final »
- **Elaboration of method sheets**, detailing the application conditions, advantages and drawbacks, limits, and giving application examples for each method: « ScoreLCA_Matrice des méthodes_20151102_final »

Those documents were given to ScoreLCA members at the .pdf format for the procedure and .xlsx format for the method sheets and choice matrices to enable future evolution of the documents.

Here is in detail each of those steps:

- **Procedure sheet**

  The procedure sheet aims at giving to the practitioners, before they start their LCA study, a global overview on the relevant steps in terms of missing data, and for each step to indicate which are the appropriate methods, and the essential questions to answer in order to help with this choice.

  Those questions are:

  - **LCA upstream phase**
    - Is there a reference document?
    - Has the firm already conducted similar studies?
    - Does the LCA include a comparison with another system?
  - **LCA phase**
    - Does the firm have access to the product?
    - What are the available LCI data sets?
    - What are the available LCA tools?
  - **LCA downstream phase**
    - Is it necessary to perform a new iteration of the LCA?

- **Method choice matrices**

  During the methods analysis, several criteria have been identified as decisive for the selection and application of methods.

  So the practitioners will be able to refer to those matrices in order to determine what the most relevant methods are, depending on their constraints.

Other matrices are also available, concerning the selection of methods depending on the kind of need (identification of missing data, determination of their influence, or missing data substitution), and kind of missing data (collection data or inventory data).

- **Method sheets**

  Once the relevant methods selected, the practitioners will be able to use the method sheets which detail the information related to each identified method and complete them with examples and explications. 20 sheets have been created, covering all methods identified.

To conclude, methodological evolutions (reference documents, standards and regulations) are in progress, mainly concerning three documents:

- The ILCD handbook is to evolve in order to facilitate the entry-level conformity, mainly to facilitate the integration of those rules within the PEF methodology.
- The PEF / OEF methodological guide is still a work in progress. It should evolve in 2016 based on the feedbacks from the environmental labelling experimentation pilots.
- In France, a DHUP WG in the building sector works on the missing data question.

Moreover, each sector, each firm has its own stake and problematics, therefore a different use of the reference documents, exigencies and methods.

It is essential that each practitioner appropriates and capitalizes their own methods and exigencies database, determined from the ones identified in the report.

This appropriation can be performed through the realization of several life cycle assessments, enabling the confrontation between those tools and the firm specificities, but also to mutualize the efforts by creating a common knowledge database.

In order to allow this evolution and appropriation, the matrices and method sheets have been delivered in French in an Excel format to be modified easily.

To go further, it is important to gather and analyze feedbacks from the different stakeholders (practitioners, verifiers, study users), as well as the methods and exigencies answers to studies objectives. This could provide a clear synthesis to be used as a basis for the development or update of future reference documents.

Finally, it appeared that the missing data notion is often mixed with the data quality notion: in the definition of missing data itself, in the choice of substitution data or in the balance between completeness and precision of an LCA study. Another study commissioned by ScoreLCA concerning the different sources of uncertainties is an interesting complement to the missing data study, and enables to have a larger vision on the data selection problematic in LCA.