TALLINN UNIVERSITY OF TECHNOLOGY

Faculty of Information Technology

Department of Computer Engineering

IAG70LT

Kowshik Krishnegowda A132078

INTERNET ACCESSIBLE REFERENCE SYSTEM FOR BACKGROUND LCA DATA

Master's Thesis

Supervisor: Vladimir Viies

Supervisor: Clemens Düpmeier

Co-Supervisor: Taavi Valdlo

Tallinn 2016

Author's Declaration

I hereby certify that I am the solo author of this thesis. All used materials, references to the literature and the work of others have been referred properly. The thesis has not been presented for anywhere else.

Author: Kowshik Krishnegowda

Date:

Abstract

The main aim is to provide approaches to the reference system for accessing the LCA data. Analysing the current problems that are faced in reference systems with different tools and even recognizing the minor problems that leads to the malfunctioning of the software. The analysed problems are recorded and a deep research is carried on. Thus finally providing an improvised solution on the problem. The paper also concentrates more about the issues on Data Handling and Data Exchanges between the large repositories [ILCD or ELCD]. Improper references are one of the major issues for causing the replication of data. In Europe and all over the world, people use different tools which may reside on different data formats as per to the convenience of the LCA modeller. Thus it is a huge problem of Nomenclature and accessing of the Dataset when they are imported in different parts of the world for their organization use or other. Finally, we aim to find out the other possible ways to suppress these problems in the near future and provide a better reference system for accessing the LCA data.

Keyword: LCA, Data Harmonization, ILCD Data format, Datasets, Soda4LCA [service-oriented database application], References to Data sets.

Abbreviations

LCA   Life Cycle Assessment

LCI   Life Cycle Inventory

SODA4LCA   Service Oriented Database Application for LCA

ILCD   International Life Cycle Database

ELCD   European Life Cycle Database

LCIA   Life Cycle Impact Assessment

REST [API]   Representational State Transfer

Table of Contents

List of Figures

List of Tables

Introduction

Life-Cycle Assessment (LCA) is a method to access the different environmental aspects associated with each produced product during the lifespan from creation to dispose. They involve different interlinked stages of a product from the extraction of natural resource. To implement the application of LCA, huge amount of data has to be managed and data resources has to be clearly organized such as ILCD or ELCD data repositories. LCA has a wide application in Industrial sectors regarding environmental impacts and building the production process economically.

Soda4LCA is a flexible database application for LCA data that are in ILCD data format currently. The software has a web-based user interface that can be accessed with any web user as well as a REST service interface. It can also be used by the software applications and other soda4lca database instances across the internet.

To communicate between these LCA users, large amount of data has to be transferred between each other in different sectors called as datasets. These datasets carry the information regarding each process of Manufacturing and handling the waste products emitted into environment.

Different users use different LCA tools and different reference system, datasets are linked each other according to theses reference system which may vary from one user to another.

Reference System may vary from one tool to another, which may cause the problem for the referencing issues probably with the data sets. While importing or exporting a dataset the data can be easily gathered but the problem the reference system of the tools to the end user may vary. Thus we intend to provide a generalized format for the environment to convert the data into a pre-defined reference system that will be easier to gather the information regarding the reference system of transferring data. *''Main aim of the reference system to normalize the pattern of references between the inter-linked data sets while they are used from one tool to another ''.*

The basic idea is to simplify the task of data access for the LCA users communicating over soda4lca. The objective is to keep the datasets in a single standardized reference system [ILCD] and converting the datasets to the user requirements formats while communicating between different users working on different reference systems. More detailed explanation and graphical view of the approach is provided in the further chapter.

1. Basic Concepts of LCA

Basic theoretical points about LCA and the tools working in LCA is required to understand the concept of the thesis.

1.1 Life Cycle Assessment

Life Cycle Assessment, is a technique for accessing the Environmental aspects and potential Impacts from the product generation. The LCA techniques gives an overall view for the production site form the inputs required during the production, the overall procedure for production explaining different steps [different flows which will be explained later in the part] and finally the END products that has been generated.

The input needs to be sufficient or long-lasting for this kind of production and must not run out of the resources. Thus the concept of recycling has been mainly concentrated to avoid this problem in the LCA techniques, so natural environment could be safe without any potential wastes from the production system.

LCA captures the complete lifecycle of the system being analyzed: from extraction of resources, through production, use and recycling, up to the disposal of remaining waste. The complete

walkthrough of the Life Cycle Assessment can be graphically represented in the figure 1 and explain the life-cycle steps involved during the product production.
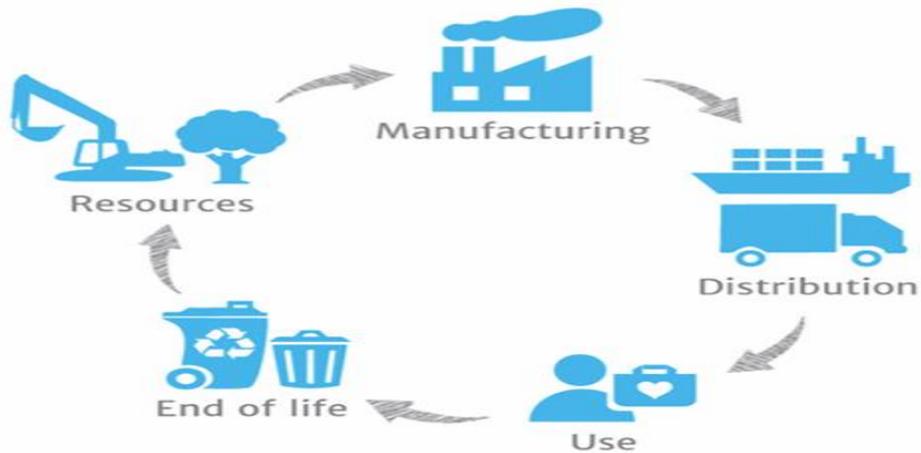
Figure 1 Concepts Illustrating LCA [1]

1.1.1 Phases of LCA

The phases explain different approaches that have been involved in LCA, during the production process can be referred in the figure 2. Each phases will be explained later on this chapter in detail [1].

1. **Goal and Scope Definition:** The produce or the service that have to assess for the production purpose it to clearly mention during the start of the service. The required level of detail has been discussed in this step.

2. **Inventory Analysis:** This lists out all the inputs and outputs for the process [Production], to be generally termed as extractions and emissions from and to the environment respectively.

3. **Impact Assessment:** Phase includes the overall calculation or results of all the stages of the process and different categories. Also calculates the potential impacts on the environment due to the certain type of emissions during the process. The results are particular to different categories which are explained as different flows later on.

4. **Interpretation:** The phase mainly concentrates on analyzing the results that are obtained by the above to phases. The interpretation includes completeness, sensitivity and consistency checks, address the uncertainty and accuracy of the results.
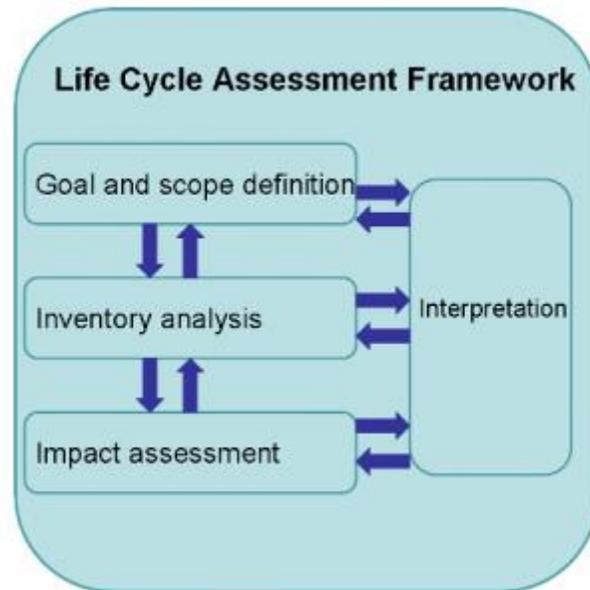
Figure 2: Phases of LCA [1]

1.1.2 Different Types of flows in LCA

LCA flows describe, the flow of materials or the physical quantity into and from the environment during the process of manufacturing a product. It represents each section how the material or product is being transformed, starting from the input until the end product is generated, and finally it is being recycled to approximately 100% or less [2]. Some part of the material which cannot be recycled is being disposed into the safe part of the environment. Figure 3 completely shows the schematic representation of the LCA during the overall production process. Each particular flows have been described briefly in the following section.
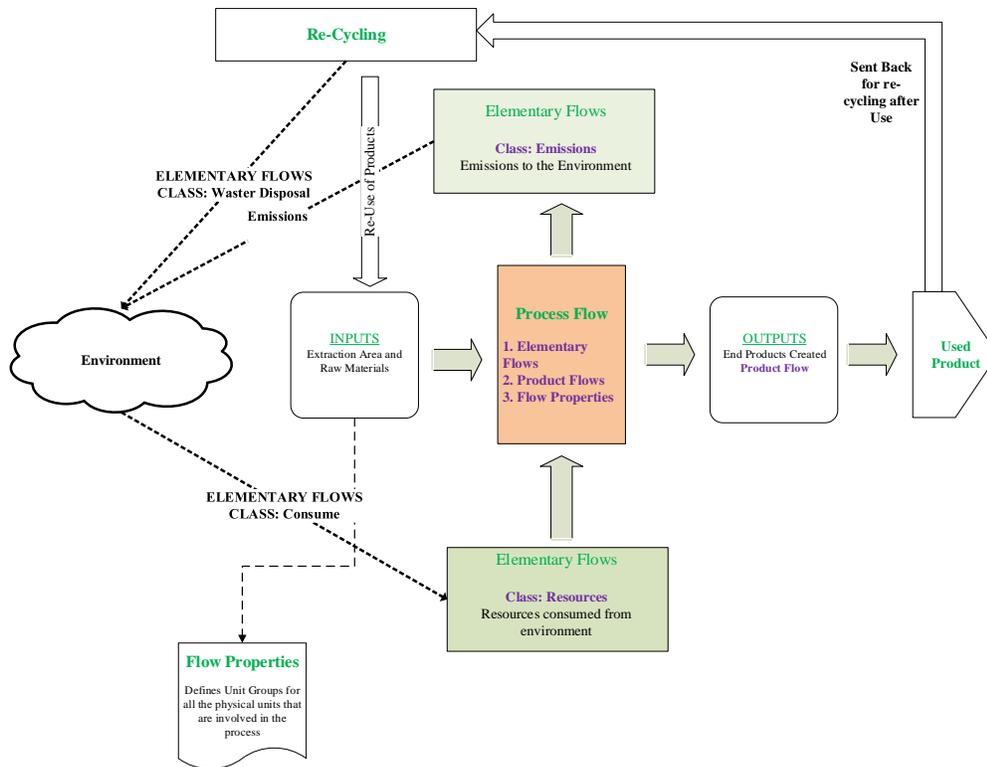
Figure 3: LCA Flows in any Production Site [2]

Process Flow: Process flow can be considered as the main system in the overall complete production site, which include series of sub-systems called flows such as Elementary flow and flow properties and finally leads into the Product flow producing an End-Product. This also includes a Unit group or termed as Flow properties, where each physical quantity used is to be measured in terms of quantity précised for the particular physical quantity is being mentioned under these properties.

They acquire the inputs [both provided raw materials and also from the Resource flow] undergo complete Manufacturing process producing an end product for the use. In the meanwhile, they also generate some physical materials that are released into the environment as Elementary flows.

Elementary Flows:

1. **Resource:** The resources that have been consumed used the environment [mainly termed as Natural resources] for the purposes of product generation. E.g. Air

2. **Emissions:** The different kinds of emissions or physical quantities that are liberated during the process flow to the environment are termed as Emissions. E.g. co2 emissions to air.

15

3. **Waste Disposal:** These are kind of emissions or waster products, after the re-cycling which must be decided by the process modeler, where the emissions are to be considered as waste or it can also be used as product to be recycled again [Product flow].

4. **Consume:** The substance which are directly consumed from the environment, without any pre-processing of the consuming substance.

Product Flow: End-Product that has been manufactured or service being delivered as a usable product or service. The output which will be used in the basic technological factor in current life for recycling. E.g. Tire or Transportation of one-ton load weighing truck between source and destination respectively.

Inputs: The provided raw materials during the initial phase of the production defined by its quantity.

Re-Cycling: After the completion of the Production process, the product is being generated. The product must be recycled to its maximum probability such a way that it can provided again used for another production process and provided as a raw material for Inputs. Since the complete recycling is not possible, some point of the recycling is once again disposed into the environment but also can be considered as the Elementary flow.

1.1.3 Role of LCA in Industries

Currently, there are different LCA tools that are used by different Industries. LCA tools helps in supporting the different phases and applications of LCA and most of them include databases. The people compute their entire manufacture process and make a clear analysis to store in the form of datasets. These datasets are uploaded into the LCA databases and used by different Manufacturers, to optimize their design and increase the efficiency. Some of the LCA tools are ccalc, Soda 4LCA, SMEs etc.

They mainly include in improving product development and improvement, identification of more sustainable products or services, eco-labelling, policy making and marketing. LCA includes a data format for import and exporting, explained in the next chapter.

1.2 ISO Standards

The main references for the paper are related to **ISO 14040:2006** "Environmental management -- Life cycle assessment -- Principles and framework", **ISO 14044:2006** "Environmental management -- Life cycle assessment -- Requirements and guidelines" and **ISO 14025:2006**"Environmental labels and declarations - Type III environmental declarations - Principles and procedures" which the standardization techniques are principled to implement on [3]. The brief summary of these references are given below:

ISO 14040:2006 describes the principles and framework for life cycle assessment (LCA) including: definition of the goal and scope of the LCA, the life cycle inventory analysis (LCI) phase, the life cycle impact assessment (LCIA) phase, the life cycle interpretation phase, reporting and critical review of the LCA, limitations of the LCA, the relationship between the LCA phases, and conditions for use of value choices and optional elements represented in figure 4.

ISO 14044:2006 specifies requirements and provides guidelines for life cycle assessment (LCA) according to scope and definitions in ISO 14040.

Figure 4: Phases of LCA according to ISO 14040:2006 [3]

These terms are based to verify or standardize the methodology for the LCA in the environment that we are working on. The data set that are used for the final calculation of the results or the Impact Assessment of the product must contain appropriate data without misleading into wrong results.

ISO 14025:2006 establishes principles and specifies procedures for issuing quantified environmental information about products, based on life-cycle data referred to as environmental declarations. A Type III environmental declaration can be described as shown in figure 5: quantified environmental data for a product with pre-set categories of parameters based on the ISO 14040 series of standards, but not excluding additional environmental information.
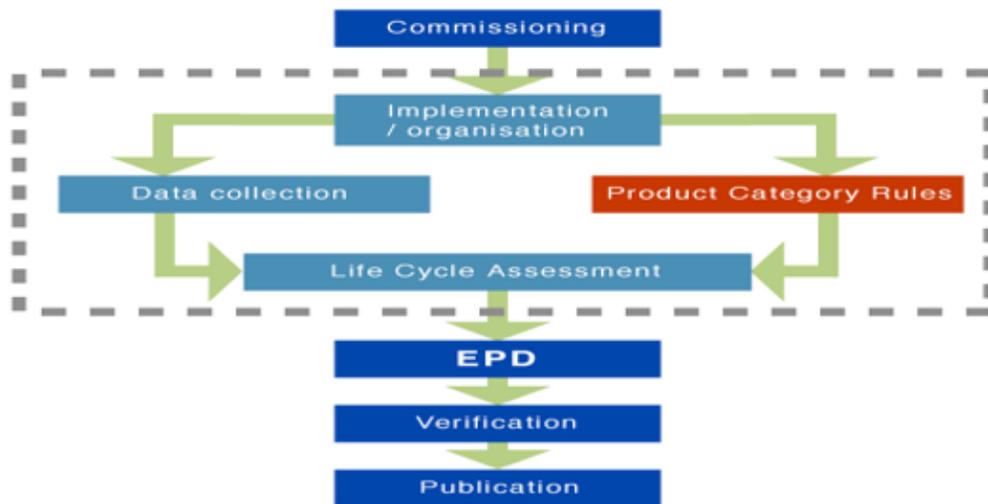
Figure 5: Stages of Product Declaration ISO 14025:2006 [3]

ISO 14040 and ISO 14025 are the main standards that have been considered in the paper for the implementation of the Reference system for Soda4LCA to access the background LCA data.

1.3 LCA Data Network

The Network of databases or the Life Cycle Data Network [LCDN] which was launched in Brussels on 6th February 2014 has mentioned some basic entry level requirements. The requirements were the latest release or the fourth phase and which consists of the "Unspecified data quality" was introduced [4]. They bring out two main reasons:

1. Making as much data available as soon as possible without compromising too much in terms of consistency and quality-assurance.

2. Providing both incentives and a clear time-plan for improving stepwise the minimum quality of the data that is available via the network.

The web-based infrastructure to ensure LCA data with easy access for all the end users. The availability of the data for the users must be quick and be able to assess their products as soon as possible. The publishing data sets must be registered with Life Cycle Data Network [LCDN] must be complaint with the ILCD Data Network Entry-Level requirements as mentioned below table 1.

Table 1: ILCD Data Network Entry-Level Requirements [4]

Compliance area	ILCD entry-level area
Documentation	• Minimum documentation extent specified • ILCD format to be used
Nomenclature	• To be compliant with 'ILCD Handbook –Nomenclature and other conventions' document (including use of ILCD reference elementary flows) • Minor deviation is allowed regarding the inclusion of commonly accepted aggregated elementary flows • Terminology use not to be enforced
Data quality	• Data quality to be stated using ISO quality criteria • Technological, geographical and time-related representativeness to be documented
Method	• ISO 14040 and ISO 14044 compliant process-based LCA • Applied modelling framework(s) and allocation/substitution approaches to be documented
Review	• Qualified reviewer' required (based on ISO 14025) • knowledge of relevant sector • knowledge of represented process or product • LCA method expertise and experience • Qualified independent external reviewer in line with ISO14044 requirements, but separate review report is not required (review documented in dataset) or: 'Qualified independent internal reviewer in line with ISO 14044) requirements, but separate review report is required (minimum review scope defined), in addition to documentation provided within data set

The datasets that have been registered or to be published in LCA data network must be processed with some of the ISO standards as described above. The format must be unified related to one of the ISO requirements standards.

1.4 Data Formats [Type of Reference Systems] in LCA

Data formats are basically it's a simple way for representation of data. LCA includes data exchange as the important aspect, thus Data Formats play a very important rule in LCA. Basically they are called as LCA data formats. The formats which is used to store the datasets in a database and software is not directly relevant.

The major types of LCA Data formats are:

1. **Ecospold Data Format:** Ecospold is a common data exchange format used for Life cycle inventory data and Life cycle impact assessment methods. Most of the LCA tools Ecospold data format interface to enable easy importing, editing and exporting of Ecospold datasets. All of the Ecospold datasets rely on the Eco invent database or repository [1].

Currently we are not working with this format in our thesis, thus much information about this data format is not provided.

Figure 6: Format Converted of Eco-Spold [5]

In the above figure 6, the OpenLCA is a tool which is trying to convert into Ecospold data format.

20

2. ILCD Data Format: This data format is one of our main concerns in our thesis and described much more detail in for-coming section. The ILCD Data Network is a web-based infrastructure allowing convenient online access to consistent and quality-assured life cycle inventory (LCI) data sets from various providers, globally.

It is an improved data format, which mainly involves the Inventory calculations. ILCD method data set specification is implemented on software and the tools will be enabled to perform the LCA operations on the available datasets in the database. The ILCD data format basically uses the ELCD database or Inventory. The graphic shown in the figure 7 also depicts ILCD data format as an option.

1.5 ILCD Data Format

Different companies or industries deploy different LCA Tools to exchange the data across the different stakeholders all over the world. Thus they need a common platform to exchange these data sets [6].

Initially before the ILCD [International Reference Life Cycle Data System] data format has been introduced, these kind of data sets have exported and imported among themselves using ELCD [European Reference Life Cycle Data System] data format. These dataset format has been used earlier to exchange their data sets for their concurrent advantages among the European databases. Even though it has been used for lot of years, it was not able to meet the needs of the developers due to the lack of different language availability, not well scalable, not able to link flow charts, pictures, complicated diagrams via electronically.

Well, later on the ILCD format was introduced in 2009 with included features that was lacking in ELCD format. The ILCD data set format was a great platform or dataset format to exchange the data [Import and Export] regarding the Assessment process with different LCA databases and different LCA tool. **The current ILCD format was further developed with so many functionalities to reach the stakeholder or developers requirements. But even though, it lacks in referencing the datasets that are used by the different LCA tool and exchanging with different LCA databases which we are going to explain in this concept in detail.**

21

1.5.1 ILCD Data Sets

Datasets are collection of data respective to the particular process. ILCD datasets are differentiated based on their functionalities and their source collection. The datasets are stored in the central location which can be shared by all the developers and used for the appropriate application. Each and every dataset have the corresponding references linking to another dataset which carries the complete information about a process while implementing in the LCA methodology [2]. Thus ILCD dataset format is made for wide compatibility and allowed for complete documentation of the datasets. ILCD datasets have been classified has following into appropriate source of information. The inter-connections and references are shown in the figure 7.

1. **Process data set:** This provides the complete information that the process is engaged with regarding that each step of the LCA method. The particular dataset has the references to all other dataset and also includes the respective flows which provides the detail information.

2. **Flow dataset:** Basically insists on all kind of elementary flows that are flowing to and from the environment. They also include the part of the inputs and outputs to the current process which may also be considered as type of the flows in the overall LCA process.

3. **Flow properties:** They consist of main physical substances that have been liberated or consumed to and from the environment in terms of their respective units. The environmental flows are clearly characterized under this sections and related data is being provided. The **unit groups** under this sections mention the measurement unit in which they are idealized for consideration for a particular kind of process.

4. **LCIA method dataset:** It's an aggregate data set which consists of all the related information regarding the methods or process involved. They also contain the data regarding the origin of the methods and detail contact set which are involved within this dataset. They represent different flows in the methods and their corresponding impacts on the environment. To be shortened as, it's a result data set containing detailed information of a process.

Figure 7: Major ILCD Data Sets [2]

5. **Contact Data set:** The contact data set provides the details regarding the administrator of the complete dataset for the person created the method. The concerned authority may be a Private Companies, Government Organization or many several other organizations.

6. **Source Data set:** They are used to classify the type of the data set that are existing currently regarding a method or a process. This may easier the fast identification for the different source classes that may have a special function and also often referenced from within the process data sets.

23

1.5.2 Core Data Sets and Core Process in ILCD

The data can be classified in to different types according to their application. Considering the working of a current Process the data which are collected directly from the supplier or the data used for the calculation on the foreground is known as Specific collected data. The process may also use the Background data or the Secondary data which is not directly collected, measured, or estimated, but rather sourced from a third-party life cycle inventory database or other source that complies with the data quality requirements of the Environmental Footprint method. These data are indirectly related for the LCA process to calculate the impact.

The Secondary data is further divided into two different types of data based on the information of the datasets they carry. These different datasets are graphically explained in figure 8:

- Product/Sector Specific Data (PSSD): it refers to data that are used for all processes that are product or organization specific (e.g. the production pulp is involved in paper products Life-Cycle);
- Cross-Sectorial Secondary Data (CSSD): it refers to data that is used for processes that are common to Life-Cycle of several kinds of products or organizations. Typical CSSD concern energy carrier production, packaging production and transportation processes.
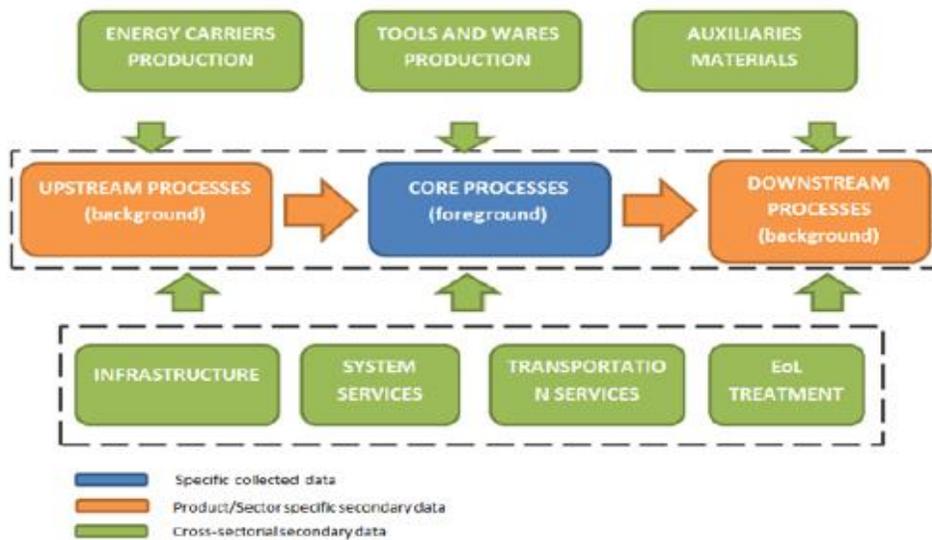
Figure 8: Relationship between core data and core process [7]

LCA covers a wide range of datasets perspective to different manner. Currently the number of datasets is reaching a very large number and causing issues in Data Harmonization. ***Problems are normally in the format of the dataset that has been imported or exported to different sectors. The paper mainly aims to overcome these kinds of referencing problems with an application called soda4LCA, where the datasets from different repository are found.***

1.6 LCA Tools to Handle LCA Datasets

LCA tools are not only used to access the repository of huge datasets but also to undergo various calculations and implementations of the datasets during the manufacturing process. They consist of broad range of features and many available databases. Software plays a very important role in categorizing different datasets of different flows into their respective streams. The impact of the overall manufacture process of the system can be calculated and used for the analysis [8].

The tools are required to build the LCA with a local repository or database, where the information regarding each step and each method are described. This can be done in a LCA software which is basically handled by an LCA tool. There are different LCA tools respective their organizations and uses for using their tool, described in the next section.

Wider applications of LCA tools:

1. Wider applications in industry, agency and research facilities doing commercial work for LCA.
2. Non-commercial Usage. [Thesis, paper, discussions etc.].
3. For Environment sustainability and Occupational safety.
4. Maintaining the Pollution constraint.

1.6.1 LCA Tools

LCA tools are for conducting LCA or for supporting the different phases and applications of LCA. Currently exists number of LCA tools from various organizations. They identically maintain their own Data repository and different syntax for their Data formats to store their datasets in the

database. Thus the person or the vendor using their application can make an easy access to their database and acquire all the required information for their production process.

The person using different LCA tools cannot handle the two databases at the same time, since the data formats and terminology for their different datasets are completely different. User cannot have an access the two different repositories respectively uniquely with a single LCA tool. Here are view examples of the LCA tool table 2 below:

Table 2: Different Types of LCA tool used currently [8]

Tool-name	Supplier	Supports LCI and/or LCIA	Supports full LCA*	Language	Main database	Special area if any	Free?
OpenLCA	Green Delta TC GmbH		Yes	English			Yes
SimaPro 7	PRé Consultants B.V.		Yes	E.g. Spanish, French, Italian, German, English	SimaPro database		No
Ecoinvent waste disposal inventory tools v1.0	Doka Life Cycle Assessments (Doka Okobilanzen)	Yes		English	Ecoinvent database	Waste management	No
GaBi 4	PE International GmbH University of Stuttgart, LBP-GaBi		Yes	English	Gabi database		No
Environmental Impact Estimator V3.0.2	Athena Sustainable Materials Institute		Yes	English	Own database	Construction industry	No

Open LCA: They provide the technical means for data exchange between the most important LCA data formats, enabling user to switch between them.

The tool converts EcoSpold01 to ILCD, EcoSpold02 to ILCD, ILCD to EcoSpold01, and ILCD to EcoSpold02. Converted files can also be used by other LCA software's, beyond openLCA. The converter runs as a stand-alone application on Windows, Mac, and Linux operating systems.

Gabi: GaBi models every element of a product or system from a life cycle perspective, equipping businesses to make the best informed decisions on the manufacture and life cycle of any product

It also provides an easily accessible and constantly refreshed content database that details the costs, energy and environmental impact of sourcing and refining every raw material or processed component of a manufactured item. In addition, it looks at the impact on the environment presenting alternative options for manufacturing, distribution, recyclability, pollution and sustainability.

SimaPro: SimaPro is most often used LCA software worldwide and their latest contributions have given great advantage towards LCA. They provide standardization and top flexibility that users can configure the way the need.

It has unique features such as parameterized modelling and interactive results analysis. It mostly includes a uniquely complete implementation of the world's leading database: Eco invent.

These are the few examples of LCA tools and exists many other, soda4LCA is one among them in which my thesis work is carried on.

2. Architecture of Soda4LCA

2.1 Working Background

To Implement an LCA application in a real world, it requires a tool called LCA tools to manipulate the data and organize them into perfect groups, to perform precise calculations to produce an efficient product in the Manufacturing process as shown in the figure 1. The different tools provide different organizational methods and procedure to perform LCA on manufacturing process. Different LCA tools have their own databases to access their datasets and their own data formats for the exchange of data between the huge data repository. LCA modelers or LCA experts use these applications to simplify their production process.

SODA4LCA was developed by Institute of Applied Informatics in Karlsruhe Institute of Technology, Germany. My topic is concerned to the data harmonization problem regarding the software soda4lca. The more detailed explanation of Soda4lca is provided in further chapter 2.

Soda4lca is a database instance which provides an access to all the different type of datasets that are provided by different LCA modelers using different tools. Each LCA user uploads their data into their repository concerned to their LCA production and soda4lca can also acts as a Web User Interface to access their database.

2.2 Soda4LCA

Currently it provides an access to different repositories but lacking in referencing systems which is a huge problem for the LCA experts while importing or exporting the data, since the datasets are in different data formats in different repositories [9].

Soda4lca is a Service-Oriented Database Application for LCA developed by Applied Computer Science [IAI] at Karlsruhe Institute of Technology for the International Life Cycle Data System [ILCD] Data Network. One of the main existing problem in LCA modelers in the modelling the data, since the datasets are required from the different database and have different nomenclature

to recognize them during the LCA studies. Soda4lca is one of the approach to solve this existing problem which depends on the RESTful [Representational State Transfer] Service API [10].

RESTful service defines a computer language- and operating system-neutral API based on Internet-conform HTTP calls and return documents in language-independent formats, like XML, clients which use the SODA4LCA service API can be written for any programming environment providing HTTP and XML functionality.

The RESTful service approach deeply allows the provision for linking between the different physical instances of the database software across all over the internet and different users providing a harmonized Network of LCA databases. The description of the node and their access to different databases across the network is schematically shown in the figure 9.
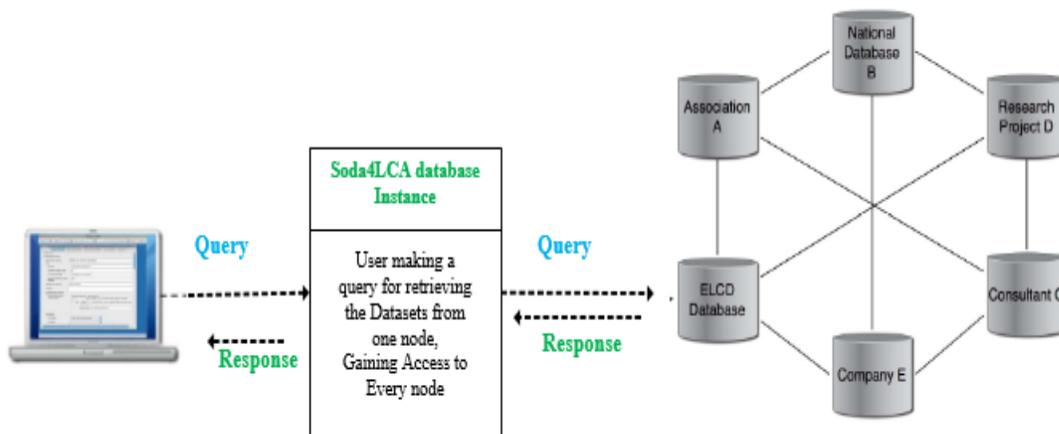
Figure 9. Data Network with different LCA Databases [9]

Soda4lca has a very wide application over web interfaces which includes a 4-layer architecture with a database as shown in figure 10, data management system, actions to perform and at the top layer with Web User interface and RESTful service. In a RESTful distributed service architecture, every service is considered to provide representations of Arbitrary abstractions of real-life objects called resources.

Soda4LCA includes functionalities for import and export, and search and retrieval of datasets. An access right management system allows administrators to define which users have what kind of access (read, read only metadata, import, overwrite) to datasets. A further dataset administration

module allows the management of datasets in a soda4LCA database instance. Figure 10 shows the schematic representation of the architecture of the Soda4LCA.
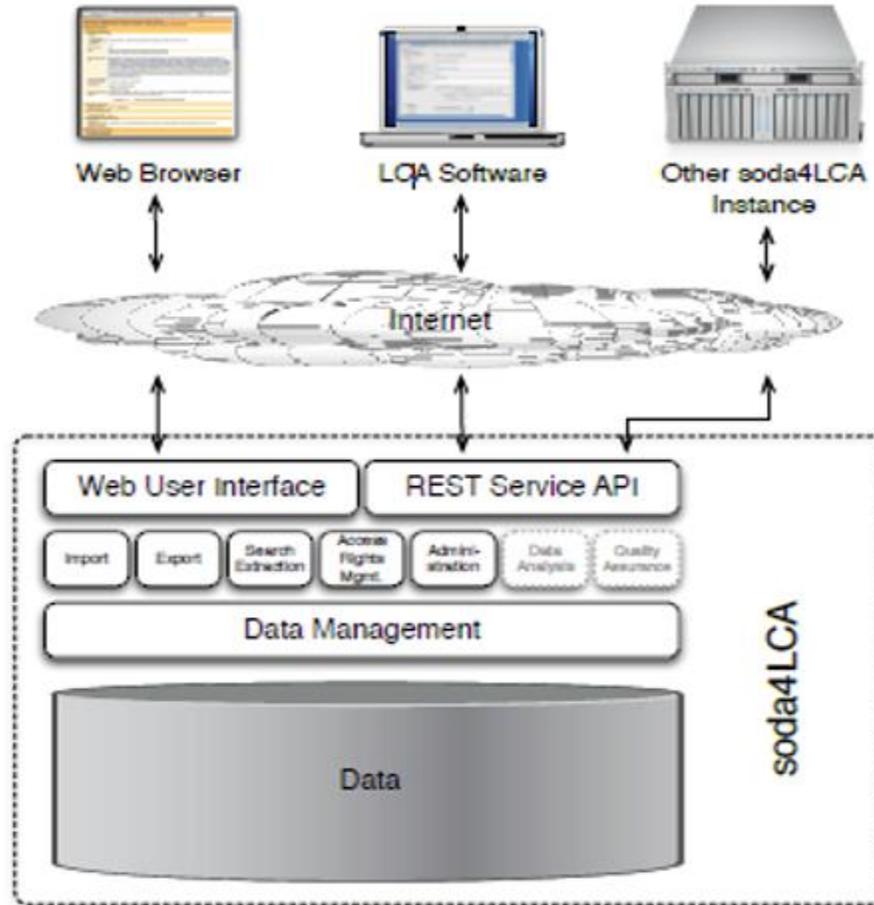
Figure 10: Architecture of Soda4LCA [9]

Thus the Soda4LCA provides a consistent access to the datasets that have stored in the central database system. Thus all the users can find their respective datasets concerned to their arbitrary node which can provide access to all the datasets in a single network from a single point. Which in case the tools also use the service API which can have an access to the complete network.

2.3 Soda4LCA Data Semantic Approach

Currently many applications of LCA [Life Cycle Assessment] or LCA tools have an availability of accessing the online datasets stored in the huge repository called as **ILCD** [International Reference Lifecycle Data System], **ELCD** [European Reference Lifecycle Database] and so on. While exchanging theses huge amount of datasets LCA users or LCA experts are facing huge harmonization problem regarding the data semantics, depending on many different specifications namely content, format and transmission of data. They can be also sub-categorized into General LCA data and Core LCA data.

For allowing easy usage of LCA data, all data used within the same tool must adhere to some common semantics and should be automatically connect able along the modelled process chains for a certain production. Therefore, LCA tools adhere to some predefined semantic model which aligns all (internal) data sets within a common reference system. Such reference systems may vary from one tool to another, which may cause problems when data are exchanged between tools and cannot automatically be transformed from one reference system to another.
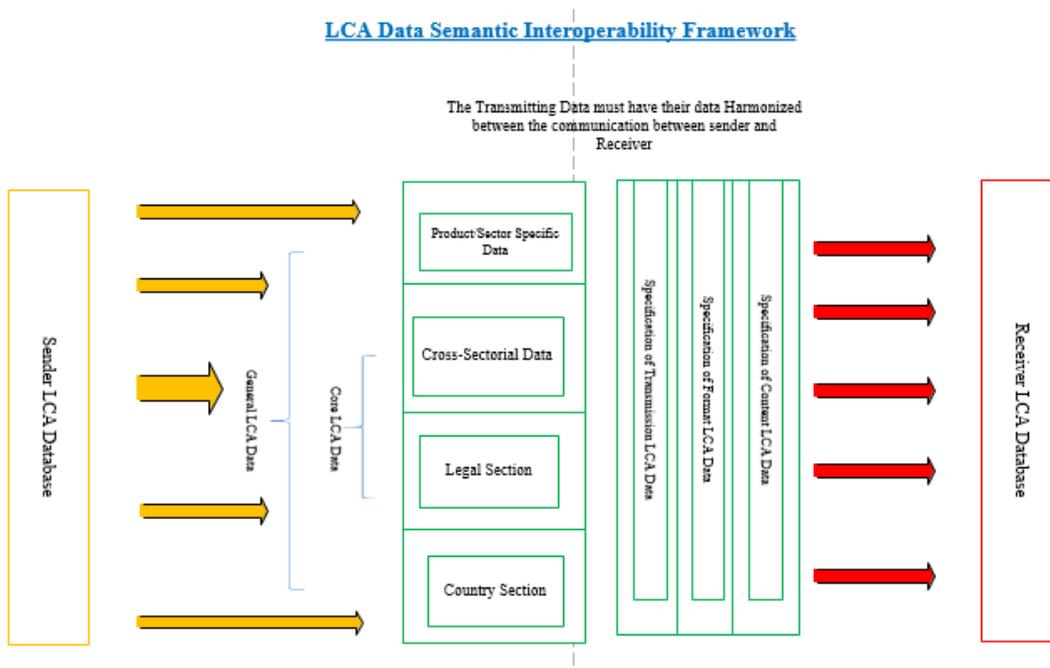
Figure 11: LCA Data Semantic Interoperability Framework

The general sector is included as the Product or Sector specific data, Country specification, transport etc. While the core sector includes cross-sectorial data, legal section, measurement units etc. Whenever there is an exchange of the data between 2 LCA experts, the transfer has to consider all these harmonization issues before importing the data into their local standardization. The graphical presentation of this problem is illustrated in the figure 11.

To overcome this kind of data harmonization problem, we are providing a reference system to exchange these LCA data between the different LCA experts. Thus the reference system concentrates on data semantics while the data is being exchanged between the two different users to avoid duplication of data or missing of data. Since, soda4LCA is a database instance, while a user accessing these datasets online, the system acts as a background reference system for the exchange of LCA data. The working environment and description of reference system is explained in the fore-coming sections.

3.4 Application of Reference System in Soda4LCA around LCA Data Access

Reference System is used to solve the data handling and data harmonization problems that are existing in the current life.

The harmonization problem is mainly concerned with the exchange of data created by independently working users that use different LCA software. LCA datasets can depend on each other and especially larger aggregated process datasets surely depend on important background data sets (such as raw materials, energy or transport), and secondly every process dataset calculates environmental flows (called elementary flows), like emissions, which relate to a certain flow substance (e.g. carbon dioxide) and compartment (e.g. flow into "air"). The harmonization problem is now concerned with the fact, that each of the above described dependencies (i.e. products flowing from one process to another or elementary flows flowing into the environment) must be resolved by "unique references" which point to a CERTAIN product or elementary flow dataset. But independent users creating datasets may use different names and different ids as identifier for their datasets resulting in non-harmonized datasets.

Soda4LCA currently doesn't provide a solution for the harmonization problem. While soda4LCA can provide LCA dataset repositories for different organizations and vendors and has data quality

mechanism to check if references from one dataset to others are not correctly resolved, it doesn't guarantee or enforce harmonization of datasets per sector. But as the number of datasets managed in soda4LCA instances grows rapidly soda4LCA functionality to support harmonization of datasets across a network of nodes are instantly needed. The work carried on the thesis is graphically represented in the following figure 12.
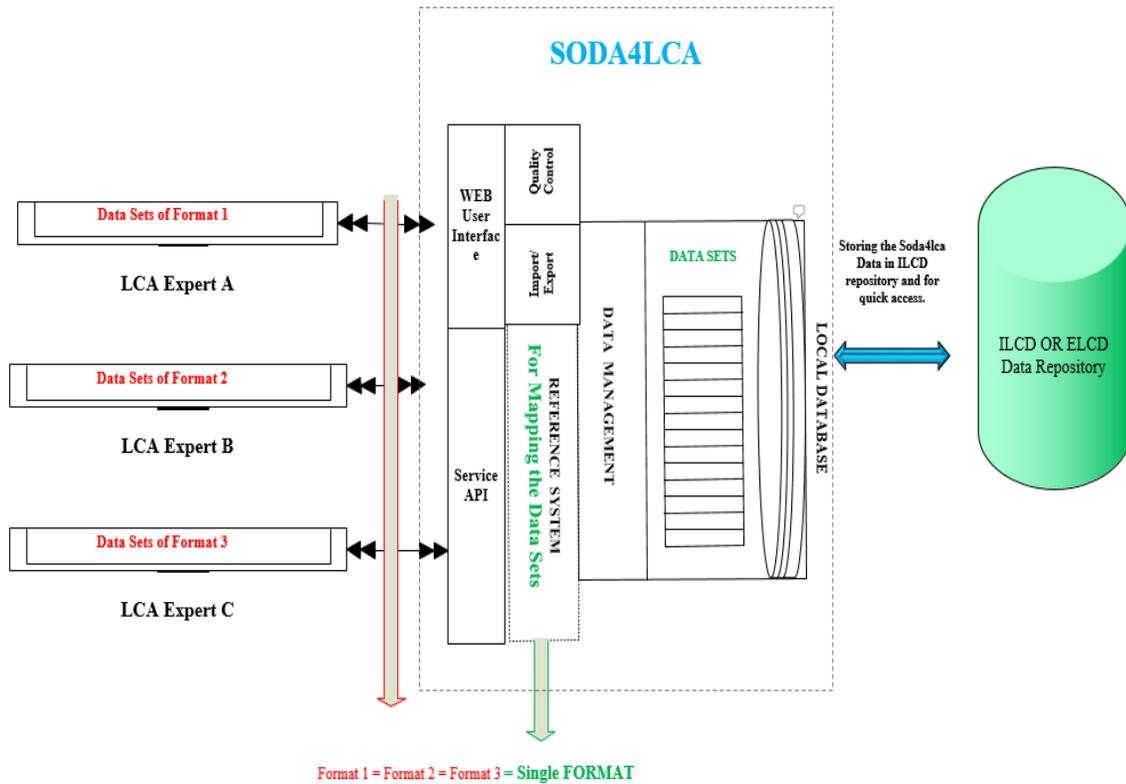
Figure 12: Reference system for SODA4CLA to access background LCA data

The above reference system clearly implies that software soda4LCA has all application for accessing different datasets from different database for the LCA modeler using soda4LCA except than that of the reference system which has to be implemented. Thesis is mainly focused on the referencing the datasets that are acquired from different users which are of different data formats.

3. Referencing Concepts for Soda4LCA

The problem of reference system may appear in different layers or to be more specific, in different perspectives while accessing the LCA data with different tools. In the current chapter, we provide the problem that can affect during data access of LCA with respect to Data Harmonization and include the solution approach to solve the recognized problem.

Limited data availability is one of the main constraints that has been noticed in existing LCA modelling projects. This mean that their already exists a data but they are not easily transferrable due to the use of different LCA tool by different LCA modellers. Since the LCA is inter-disciplinary field requires synthesizing information from a variety of discipline specific departments to produce an overall resulting output product. Often the parameters of the several departments keep on changing and there must be a measure to be taken they relied on a single system.

3.1 Connecting Process Chain problem

The connection of the datasets in an process chain has been one of the major drawback for the LCA modellers, since the different LCA modellers use different tools and the unique identification for each of the product has been changed thus ending up with two different UID's for the same product in an Organization or between inter Organization. The problem has been explained in the following scenarios:

3.1.1 Scenario 1: Handling Products of Process Chains in Single Organization

Let's consider the case in an Organization undergoing LCA production and developing a product in each different process under various departments [11]. The Process 1 utilizes some of the extracted products as inputs and with some of the predefined steps it produces an output product called COAL [example]. The information has feed into the database with an UUID defined by the modellers of Process 1 considered as in the figure 12.

The Process 2 needs to develop an output product which uses the COAL to produce its own end product. It searches for the database for the coal with an UUID of 121 which it is assigned by itself for its production. Each process providing their UUID for their physical quantities may be great harm in harmonizing the data.
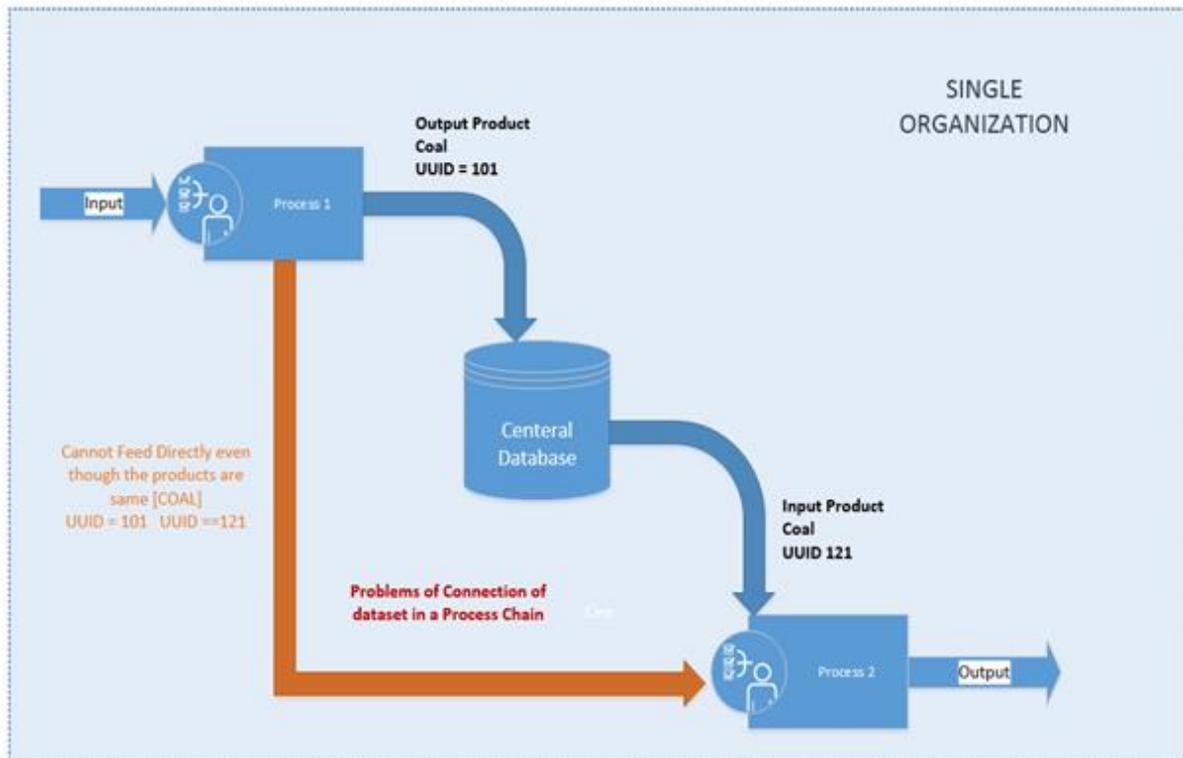
Figure 13: Problems in Connecting the Process Chains

Even though the product produced by the Process 1 and the product required by the Process 2 is the same in physical terms but they cannot be feed directly even though they are in a single organization because their UUID are different and the Processes cannot be interdependent on each other. The schematic representation of this has been showed in the figure 12. This can be one the threat when we are trying to make an access between the inter-organizational processes for trying them to relate in one single box.

3.1.2 Solution to Scenario 1 [Handling Products of Process Chains in Single Organization]:

Every organization uses their own tools to access the huge LCA data of different reference systems. Even though they access the wide range of data with their respective tool of its own feature but it also very important for the Industries to maintain proper reference system when they are using reference system for communicating between their processes as explained in the previous section 3.1.1.

In an organization, the centralized reference system can be implemented. The main task of this system is to organize the unique UUID for the datasets in the database irrespective of the UUID that has been provided by the modeler to the datasets during the process. The graphical representation of the solution is provided in the figure 13 below.
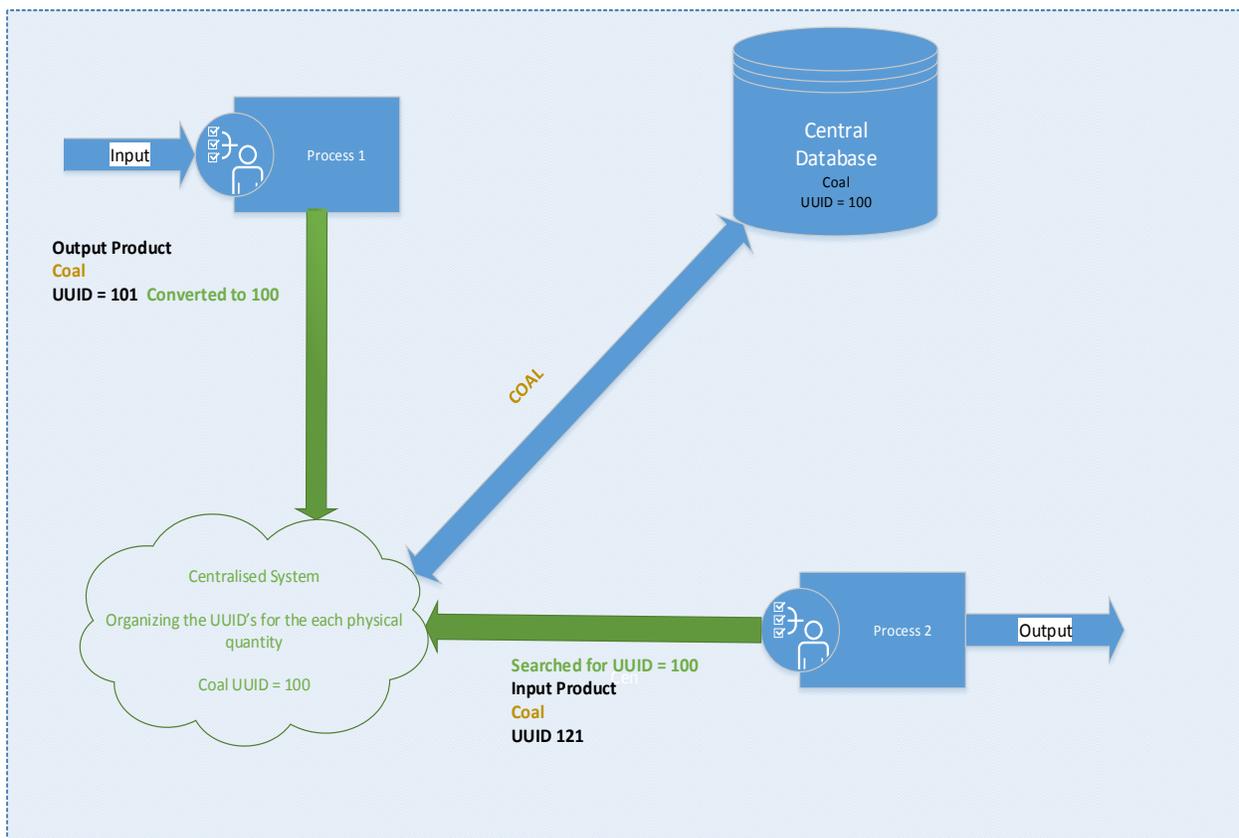
Figure 14: Centralized System for Process chains in an Organization

After the production of the process 1, the output product is created and a UUID value being assigned by the modeler. The information of the created output [coal as an example according to

the previous section] is not directed to the database for storing the information of the output dataset. The centralized system has the information [name, flow, units and etc.] of the product and assigns a unified UUID from 101[Assigned by modeler process 1] to the 100 [assigned by centralized reference system] and transfers the information to database to the storage.

The modeler in the process 2 is searching for same product as generated by the process 1 [Coal] and having its own UUID of 121 [assigned by modeler in process 2] for coal. When the modeler request the centralized system for the product coal, It checks the dataset in the database for the required product based on its name and other required information, fetches coal and redirects to the modeler. This way the UUID of the product in an organization can be unified to avoid confusion of the datasets in a database.

3.1.3 Scenario 2: Abstract Specific Product

During the process chain of a product, the production unit modelling has a defined Input source which is already mentioned during the designing of the production unit. Since there is no more option for the Modeller of the product unit to make a right choice in-between. Thus the modeller cannot improvise the current modelling techniques of the production unit in more efficient manner. If the modeller is requesting for a specific kind of Input but there is no option available to choose as per the Modeller requirement. Thus the modeller has to deal within the system boundaries and select the available resource within them.

System Boundaries: System boundaries can be set geographically and in time, and concerning the processes that are included in the system. Geographical system boundaries are important to consider, as the different lifecycle phases may occur in different regions and thus the systems/processes included (e.g. electricity production, transportation, etc.) might differ significantly between the life cycle phases [2]. The system boundary section is represented clearly in the figure 14.

Let's assume as showed in the figure 14, the process of Brick production in an Organization. Which involves a lot of the process steps for the production of brick which can be implemented in LCA and calculated the Impacts within the system boundaries. But the entire production uses the electricity which is yielding from an outside source for the production.

37
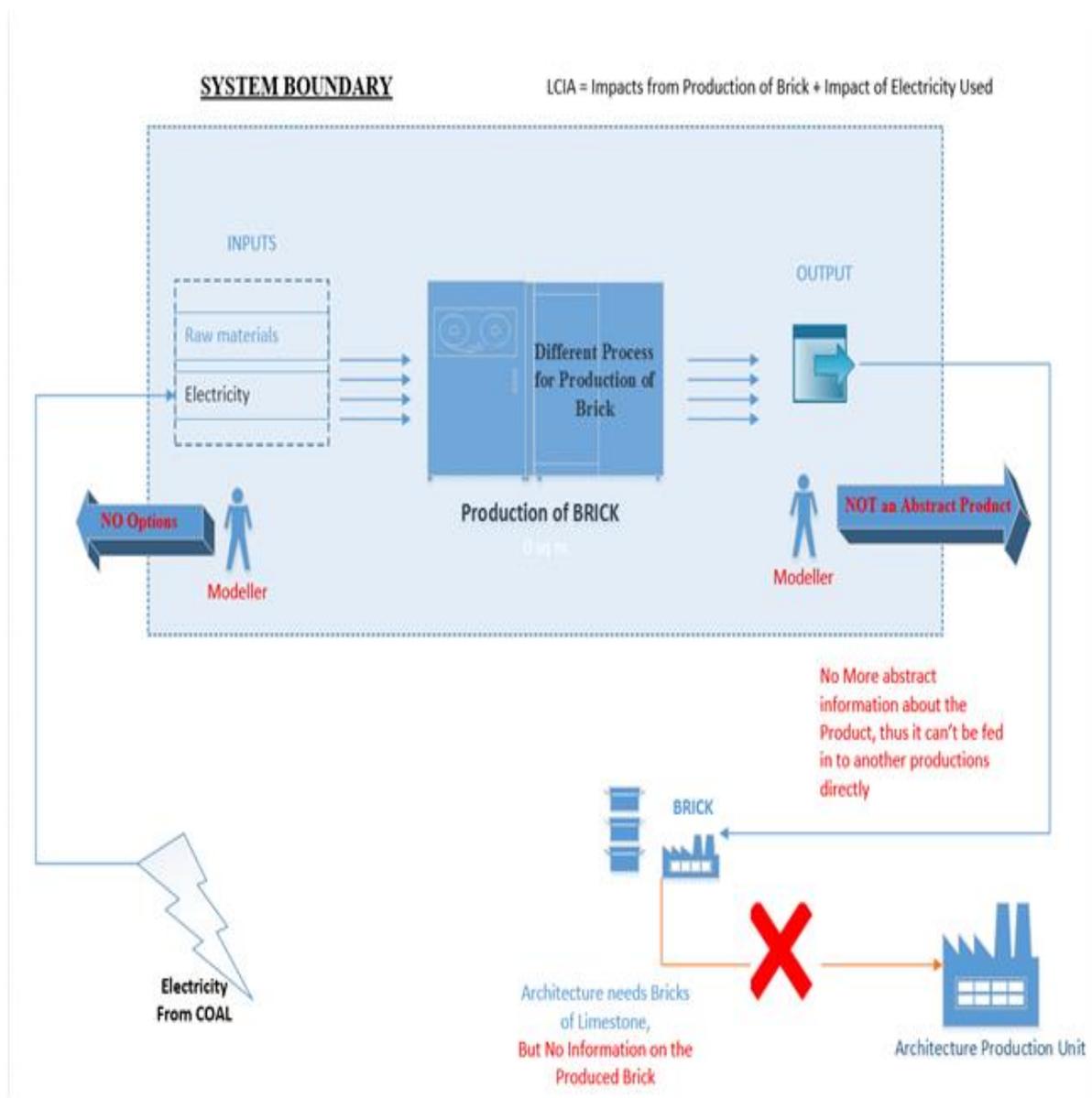
Figure 15: Problem due to lack of Abstract Specific Product

Brick Production involves lot of intermediate process within the production unit for the generation of the end-product [Brick]. But all through these intermediate process, the unit also requires electricity which do not lie within the system boundary of the Brick production. The electricity is fed by an outside source from a different unit, here in our example it's from coal. Thus the overall impact is calculated, it also involves the burden caused for the environment for producing the electricity that has been used by the production unit. Which in case, the modeller has no other options for selecting the different input source for electricity to reduce the environmental burden.

Thus the fed input source is pre-defined and no abstract information is available about the input product.

After the production of the brick, the information of the brick is updated into the database with none any other information than the brick. Even it may be a same kind of brick to be used by another production unit [Example: Architecture Production Unit] but due to lack of information on the database the modeler will not be able to decide that, Is this the same kind of brick he want to use in his production unit. Thus the modeler falls back to use these available brick in this unit. This can be a major problem stating that, even though the right product is available but due to lack of sufficient information [product specification], the products cannot be directed as the correct inputs for their further implementation in their LCA cycle.

The figure 14 clearly explain the lack of information of the product, when a product is generated in a specific production unit, during the time of uploading the information of the produced product to the database. This have to be overcome with some different approach, so that the modeller may have the more options for providing the inputs and as well as the modelling of the output product must be more precise to solve the confusion between the products. These kind of problems can be solved by providing to specific kind of references to each and every product, which are basically derived from the basic kind of a specific product.

3.1.4 Solution to Scenario 2: Abstract Specific Product

Main intension is to provide more information on the produced product or the dataset that has to be stored in the database for further access. "Abstract Specific Product means the more detailed information of the categorized dataset which shares the hierarchy of the main generalized product, thus the modeller has more option to choose the source to main the process more efficient."

Considering the problem in previous section 3.1.3, every product needs a detailed specification of themselves. Categorized products [or Datasets] need to be grouped under a generalized category where the modeller can easily access this general category of products as shown in the figure 15 and can select one of the sub-categorized product with required specification among them for the best convenience for his production process.
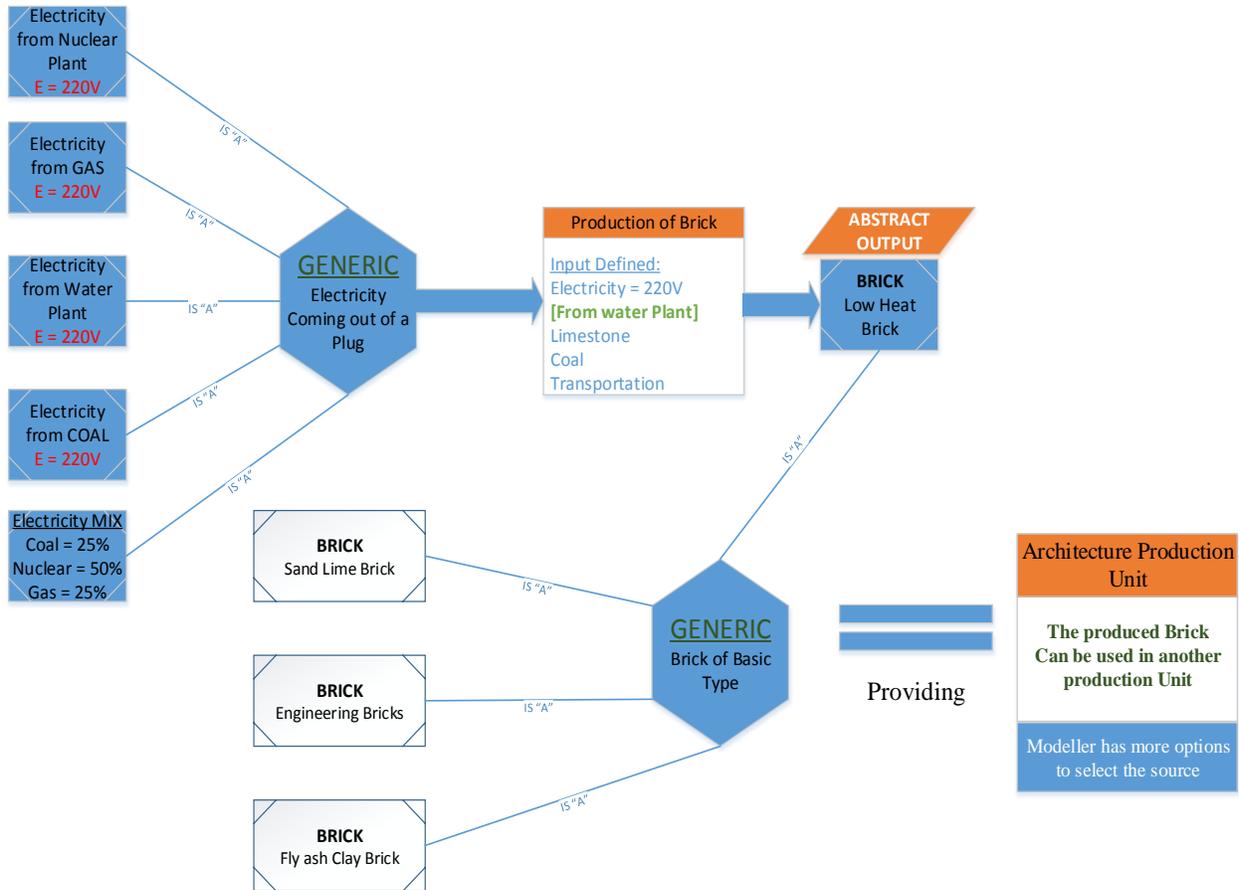
Figure 16: Specification of Abstract Specific Product

As shown in the figure 15, for the specified Brick Production unit needs an Electricity as one of the main Input source outside the system boundary from a different source. When the modeler searches for an Electricity, the database provides an availability of generalized category of product Electricity under it which describes different sub-categorized product each with different specification as shown in the figure.

Basically the modeler is looking for the Electricity that can be used which is available through a plug socket. But that is an option to look for him for the more efficient source of electricity that provides less burden to the environment, which helps his production unit while calculating the overall Impact Assessment.

The Sub-Categorized electricity of 220 V from the water plant has been chosen by the modeler to implement in his production unit of the brick. The electricity 220 V from the water plant shows an

"IS A" relationship [12] in the figure 15 which mean to say that, it is a type of Electricity still having different specifications. *"IS A" relationship describes the decoupling of product information allows for modelling hierarchies of products where of actual, vendor specific products may represent the same generic product. This relationship is expressed by a reference document as shown in the figure 16.*
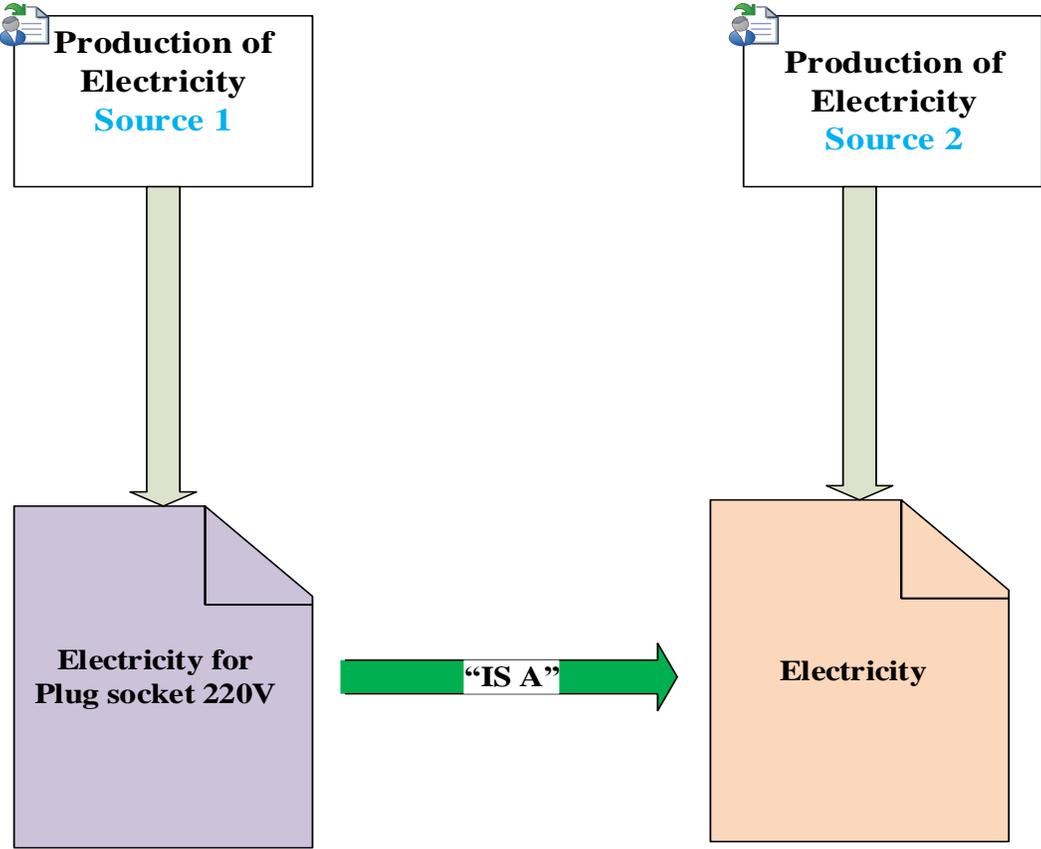
Figure 17: IS A Reference for Products [12]

Thus the generating product from the current production unit is also an abstract specific product called Brick with the specification called Low Heat Brick is linked or referenced to a Generic product called BRICK. This product is one of a type of brick but there also exists different other kinds of brick that are at abstract level specific product. In future or in any other production units [Architecture Production Unit], the brick from the current production unit which is categorized under the generic product can be used for their production purposes as input source.

3.2 References in classification of Elementary flows

One of the major confusion in LCA Analysis is the flow of the emissions into their asserted compartments. They mainly represent the flow identifying information which is totally necessary in all the processes. The ILCD has a clear classification of all the substances type and all the flow identifying classification. The Soda4LCA has an access to the ILCD database but not for the classification or categorization of the different type of substances and flows for different compartments in LCA.

During a session for processing of any manufacturing unit, the modeller using Soda4LCA software can access the ILCD database but could not make a direct access with the categorization of ILCD database. Which in turn results that, Soda4LCA cannot classify or manage or assert the flows in to their respective Elementary flows on its own, but has to make request for ILCD database to notify the classification of the current flows of the manufacturing unit. Then the resultant flows have been assigned to their particular elementary flows and to their respective compartments.

This shows the soda4LCA cannot identify the compartments [particular level where the emission is made] and their references when they have been exchanged between from LCA tool to another. When the datasets have been exchanged from one tool to another, the flows plays a major role in calculating the substantial impact assessment of the emission on the environment. During process of production each product is associated with their environmental burdens, thus confusion in these elementary flows play a huge impact.

Below figure 17 clearly representation of the classification or the categorization of all the Elementary flows and included other Elementary flows.

Different
Compartment
s at each level

From water

From AIR

From SOIL

Resource

Different
Compartment
s at each level

To water

To AIR

To SOIL

Emissions

Land Usage

Minerals from
Ground
Land
Transformation

All kinds of
occupation

Land
Usage

1. Long Term
Emissions

2. The Emissions
that are
continuously
made for long
duration in
future

3. These kind of
emissions are not
termed under
any categories
and even though
they do not
belong to.

Elementary Flows

Other Elementary
Flows

CALSSIFICATION OR CATEGORIZATION

ILCD

NO Access for
SODA4LCA

ILCD Database

NO Access for
SODA4LCA

Soda4LCA Application

Source: Info from ILCD Handbook

Figure 18: Classification of ILCD Elementary Flows [2]

To overcome this problem, we are trying to adapt the complete ILCD format reference system for all different kind of flows in soda4LCA. This avoids the confusion between 'n' number elementary flows and to their compartments. The problem is also concerned with product flow as shown in figure 17. As explained in our reference system, when a dataset has been transferred from one tool to another, the reference system is based according to ILCD reference system.

Next chapter clearly describes the purpose of reference system and the way it is managed by soda4LCA among different tools.

3.3 Reference Flow Issues for Soda4LCA Users while communicating between different LCA Tools

A reference flow is a quantified amount of the product(s), including product parts, necessary for a specific product system to deliver the performance described by the functional unit. The purpose of the reference flows is to translate the abstract functional unit into specific product flows for each of the compared systems, so that product alternatives are compared on an equivalent basis, reflecting the actual consequences of the potential product substitution.

Soda4LCA is tool used to collect datasets from different databases, where the users can access under a single instance. Users try to import datasets from different databases via soda4LCA and make a prior calculations on their host system, which do not contain the same reference system that of the database that they are trying to download form.

3.3.1 Problems regarding Different Reference Flows and their connectivity

Let's assume a user tries to download a complete process dataset from ELDC database having ILCD reference system. In case of user is using an Eco-Invent database working with EcoSpold reference system. All the datasets that are imported from ELCD is linked according to ILCD reference system to their respective flows and their connecting datasets. They are is high chance of risk that, imported datasets has many references to other flows or products may be missed or wrongly refereed to other datasets or the reference can missed too in some cases.

The process dataset will be Hard Coal which consists of several Elementary flows in input and output sections. In this scenario we are considering one of the product flow from outputs called Hard coal, at consumer EU-27 as shown in the figure 18.

Outputs		
Type Of Flow	**Classification**	**Flow**
Product flow	Valuable substances / Energy carrier / Fuels / Hard coal products / Hard coal, at consumer	Hard coal, at consumer EU-27

Figure 19: Product flow from the process Hard Coal

Hard coal, at consumer EU-27 product flow is one of a dataset classified into outputs in a collection of huge process dataset called HARD COAL. The current product flow possess a unique ID of **1bf85d1d-0b66-4476-99f9-2d69e0b019b3** concerned only to this product and also has a reference to another set of datasets in the flow column in the figure 18 and clicking on that show in the given UID in a xml format. When you open the references for Hard coal, at consumer EU-27 is looks like below figure 19.

Figure 20: References for Hard Coal, at consumer EU-27 [5]

Figure 19 clearly indicates it further has many references like reference flow property, Compliance and Dataset formats. When we try to access this individual references they point to their respective dataset with their unique identification. The UID for the described reference dataset as follows:

Table 3: UID's for the Reference datasets under Hard Coal, at consumer EU-27 product flow

Reference Flow Property	93a60a56-a3c8-11da-a746-0800200b9a66
Compliance Declarations	88d4f8d9-60f9-43d1-9ea3-329c10d7d727_01.00.000
Dataset Formats	a97a0155-0234-4b87-b4ce-a45da52f2a40

This way of Reference System Architecture is present in each and every flows or products of any bulk collection of Process datasets. As we considered in the previous scenario it is very important to handle these kind of references for bulk amount data and when they imported from tool to another which are different in their database reference system, there is high risk of losing the data or issues with connecting datasets. The particular UID for the connecting reference property may point to Null, where the importing user may not find any information regarding the particular dataset after importing for the unique UID given by the source. There is also a chance that one of the units or relating may do not exists at all in the importing database reference system.

Figure 20 clearly shows that the problems in references for the importing datasets, when they are imported from one references system to another [between different tools]. We have considered one of the dataset as *product flow* to show the link between its references to show how they are connected between each other as explained in the previous content and graphically represented in the figure 20. These problem mainly causes due to architecture of the references system in different database accordingly. When the user is trying to import the complete process datasets he imports it along with the reference system that is associated with the source [ILCD Reference system] into his database [EcoSpold] reference system, which has different method of organization of flows and connecting datasets that may or may not exist in ILCD reference system. Thus the importing user database reference system end up in allocating the improper UID and reference flows for the amount of datasets that are imported which are not synchronized as per to the EcoSpold database reference system.
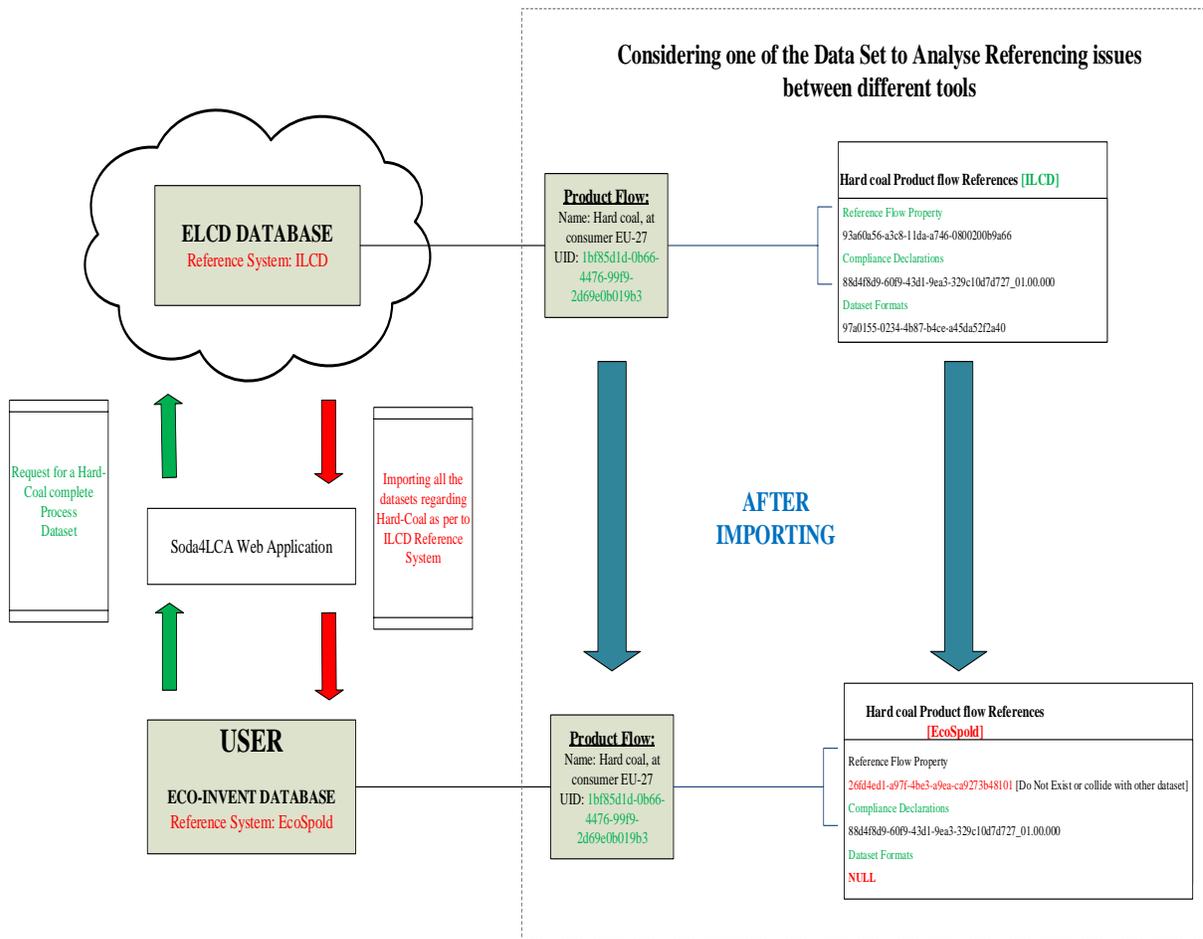
Figure 21: Referencing Issues of datasets with UID when imported from one tool to another

Reference system introduced in soda4LCA helps in converting this unsynchronized datasets into a standard form and then allows the user to import the dataset from different tool. Therefore which the imported datasets and their references to their corresponding flows are also found in the importing reference system with appropriate UID and will not lead to any loss of data.

3.3.2 Referencing Issues regarding Compartment

The problem cannot be only conceptualized to flows and its references but also stays in the area of Nomenclature and compartment for the Elementary flows. The compartments play a major role, since the number of elementary flows in each process are in huge number and the location where these flows have been liberated is a main part during the Life Cycle Assessment. The calculations of the elementary flows of any process in LCA is numerically very precise and they are very

48

important during the Impact Assessment calculation of the process involved in LCA. These impacts indicates the threat that any production unit is causing to the environment and to be reduced as much as possible.

In each Database reference system, the compartments have been named accordingly to their convenience and part of the environment [sub-compartments] where actually elementary flow have been liberated. Figure 21 shows a brief overview of the compartments for the elementary flows in any database reference system and the threats caused by them to environment.
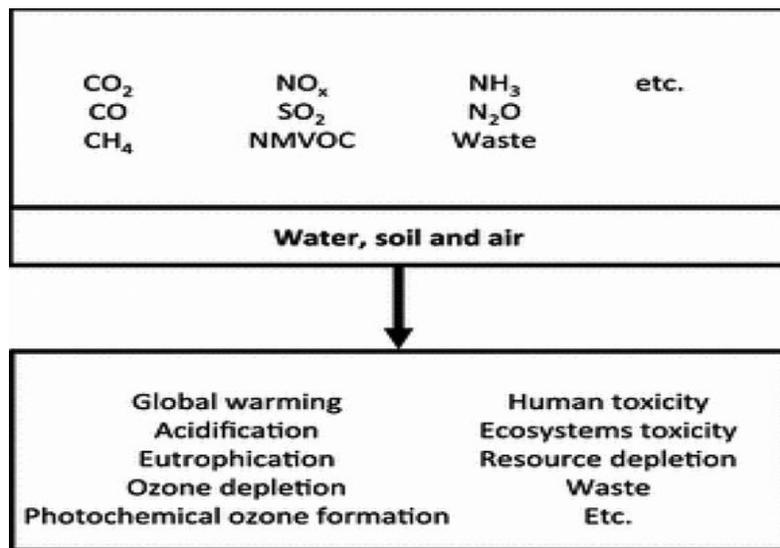
Figure 22: Compartments of LCA database reference system [8]

The compartments that have been mentioned in one reference system of the database may be not available in the receiving database reference system or also there is a risk that the compartment means the same in both reference system but the way they have named the compartments may be different which leads into a major confusion.

For example, let's assume the same scenario as we tried to explain the problems with reference flows with LCA datasets

Let's consider one of the flow from ILCD reference system as "acetic acid". User is trying to import these datasets of different flows into Eco-Invent database having EcoSpold reference system.

In **ILCD reference system**, under outputs the **acetic acid** Elementary flow is being made an emission to air is the main classification but which specifically described the sub-compartment or sub-classification upper stratosphere as shown in table 4. This shows emission must be made to the particular compartment as suggested [13].

The User downloads these datasets regarding elementary flows and saves in his Eco-Invent database. During the assignment of the flows in EcoSpold reference system, the EcoSpold tries to allocate **acetic acid** to the classification as imported. *But the EcoSpold reference system do not contain the sub-compartment Upper Stratosphere but which in case has only contains the sub compartment stratosphere as shown in the table 5.* The user is not able to decide, the assigned compartment is specifically correct or it has to be generalized into a single compartment. Due to the improper assignment of the compartment, this may result in wrong results while calculating the overall impact assessment of the product.

Table 4: Classification in ILCD Reference System

Type of Flow	Classification	Flow
Elementary Flow	Emissions/Emissions to air/Emissions to air, upper stratosphere	Acetic acid

Table 5: Classification in EcoSpold Reference System

Type of Flow	Classification	Flow
Elementary Flow	Emissions/Emissions to air/Emissions to air, stratosphere	Acetic acid

As shown in the figure 21, after importing the dataset, EcoSpold has the compartment call AIR but don not has the similar sub-compartments as that of the ILCD reference system. After importing the reference system is confused to allocate the elementary flow to the compartment which actually do not exists in the reference system.

50

Considering one of the Elementary Flow to Analyse Referencing issues between different tools

Elementary Flow:
Name: Acetic Acid
Classification:
Emission to air/
Emission to upper
Stratosphere

Compartment
AIR

Sub-Compartments

Unspecified
Upper Stratosphere
Lower Stratosphere
Air, Low density Population
Near the Ground

AFTER IMPORTING

Elementary Flow:
Name: Acetic Acid
Classification:
Emission to air/
Emission to upper
Stratosphere

Compartment
AIR

Sub-Compartments

Unspecified
Stratosphere [No Specific Sub-Compartment]
Lower Atmosphere
Upper Atmosphere
Air, Low density Population

Figure 23: Compartment Referencing Problem after Importing

In the figure 21, the transaction between the user and the ELCD database is similar to the figure 20. Figure 21 clearly explains the referencing issues between the compartments. These problems can be solved in soda4LCA by providing a mapping table that assigns the compartments and its sub-compartments to the standard format before importing or exporting with database.

3.3.3 Referencing Issues in similar Scenarios

Above explained referencing problems are not only in case of compartments or UID's but they can occasionally occur in different part of the datasets, some of them are explained below:

1. References with Unit of measurements: Measuring units can be different in different reference system based on their convenience. If the export unit measures the length in

meters and the importing system measures in centimeters, then before the exchange of the datasets, the units must be converted into appropriate values otherwise the calculation of the impacts may end in wrong results. This conversion is standardized into one of the General system in soda4lca before transaction has happened between soda4LCA users.

2. References with Database properties: The database in which the entire reference architecture is based on, has some pre-defined database properties which they do not share with the other databases. Even these properties play a major role in assigning to the appropriate datasets. Thus, the properties of the database has to set or adjusted as per according to importing database reference system.

3. References with flows: As explained in the number of previous examples, assignment of flows is a major problem. Let's assume user a references system has set the radioactive waste is mentioned as "Product Flow" but for the user B importing theses datasets of different reference system as assigned this as "Elementary flow", thus ending up with a major contradiction and resulting of fault analysis. Even though this problem is a rare case scenario but there chances of happening in totally difference reference system.

4. Implementation of Reference System in Soda4LCA

Soda4LCA acts as an interface or database instance between the different users which are connected through their respective nodes to all the users through a single data network. The current architecture of soda4lCA imports the datasets from different user having reference system A into the importing user of reference system B as it is, without any changes or assignment of the default values. This may end up in having duplication of data and missing of huge amount of datasets which are not synchronized between the importing and exporting database systems. As previously explained the architecture we are trying to implement currently is as shown in the figure 22.
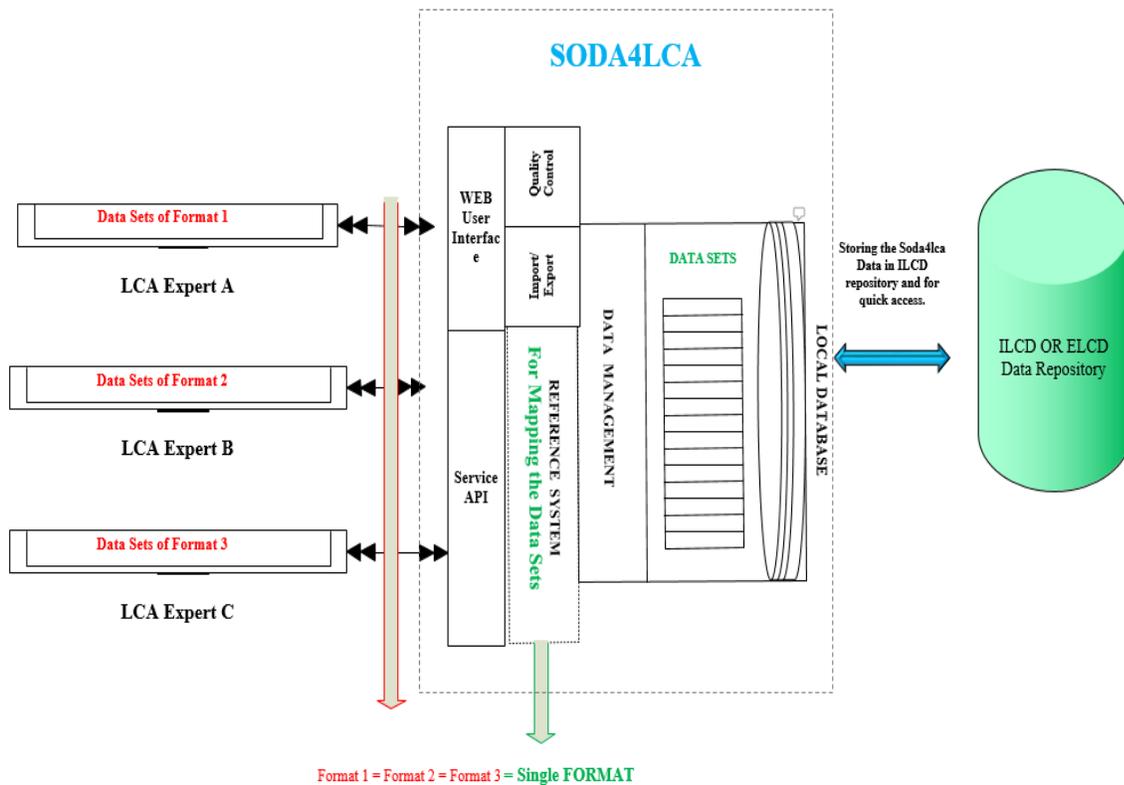
Figure 24: Reference System Implementation

Reference system functionality that has included in the soda4lca has to convert the importing or exporting formats or the reference system into a predefined reference system, in our case we are trying to convert into ILCD reference system.

*"Whenever a user is trying to communicate between the datasets all across the worlds in LCA data network through soda4lca, the **functionality of the reference system has to be implemented before the exchange of datasets has been occurred.**"*

Main intension of reference system in soda4LCA is to Integrating datasets with multiple reference system or Data formats that the database they have concerned respectively. This is important to make the different databases with different reference system interoperable [14].

Popular reference system or formats from different database like EcoSpold, EcoSpold2, Gabi, SimaPro, Open LCA and etc., has been stored in the soda4lca database. All these reference system possess similar architecture rather than some rather than naming conventions, analysis of elementary flows, unit etc. Formats from these database have been synchronized in soda4lca and stored in the soda4lca database. The similar model approaches for reference system from popular database have been reflected in soda4lca, so that it may facilitate the conversion of reference system for the exchange of datasets.

Whenever any exchange of datasets has been attempted through soda4lca web interface, Reference system in soda4lca has to convert the reference system of the source user as per required to the destination reference system and the same way vice versa. The detailed procedure have been explained the number of steps in Figure 24.

Below describes the series of steps that clearly depicts the procedure of conversion of reference system into desired database in soda4lca.

Step 1: Request from any LCA user

Request from any LCA user are sent through soda4lca web interface either to import or for the exporting of the datasets from any user using different kind of reference system. Assuming the user has a request for import the dataset from Eco-Spold 1 or 2 working in ILCD reference system

in his environment and another user having EcoSpold reference system making a call for exporting the datasets [The map for the both of these reference system must already exists in soda4lca database].

The request may be to save the database on the server or to download them. Reference System which are pre-synchronized with the soda4lca database can be converted between inter-formats, but the request by the user for the reference system which do not exists in the database is compared with the ILCD reference system as default and formats are aligned according to the ILCD reference system.

Step2: Checking for the format or reference system

Checking for the reference system used by the user in his environment. After checking the environment, soda4LCA needs to allocate the particular reference system for the user friendly environment for synchronization, to map the reference system into appropriate destination. This facilitates the comparison of all the datasets and their related references with ILCD as a default reference system.

The work of this system is to check the type of reference system that is used by the user and reference system of the datasets he is trying to import via soda4lca. If the datasets lie in the boundary if same reference system, then the datasets are directly transmitted between the user and database which directly leads into step 5 in the figure 23, otherwise the datasets have to compared and converted into the appropriate format used by the user.

Step3: Comparison of format or reference system by the user and Datasets to be imported or exported

The identified reference system has been selected from the soda4lca database and preferred reference system has set to be ILCD as an example took previously in step1. When the user is trying to import the datasets from another reference system of a different database, then the comparator request for a conversion which do not lead to direct exchange of datasets, since the

55

dataset formats have been not matching between the user environments and importing datasets reference system.

The *dataset format* of the dataset to be imported by the user has loaded into comparator for the conversion and comparison as shown in the figure 23. At this stage, the final conversion format of
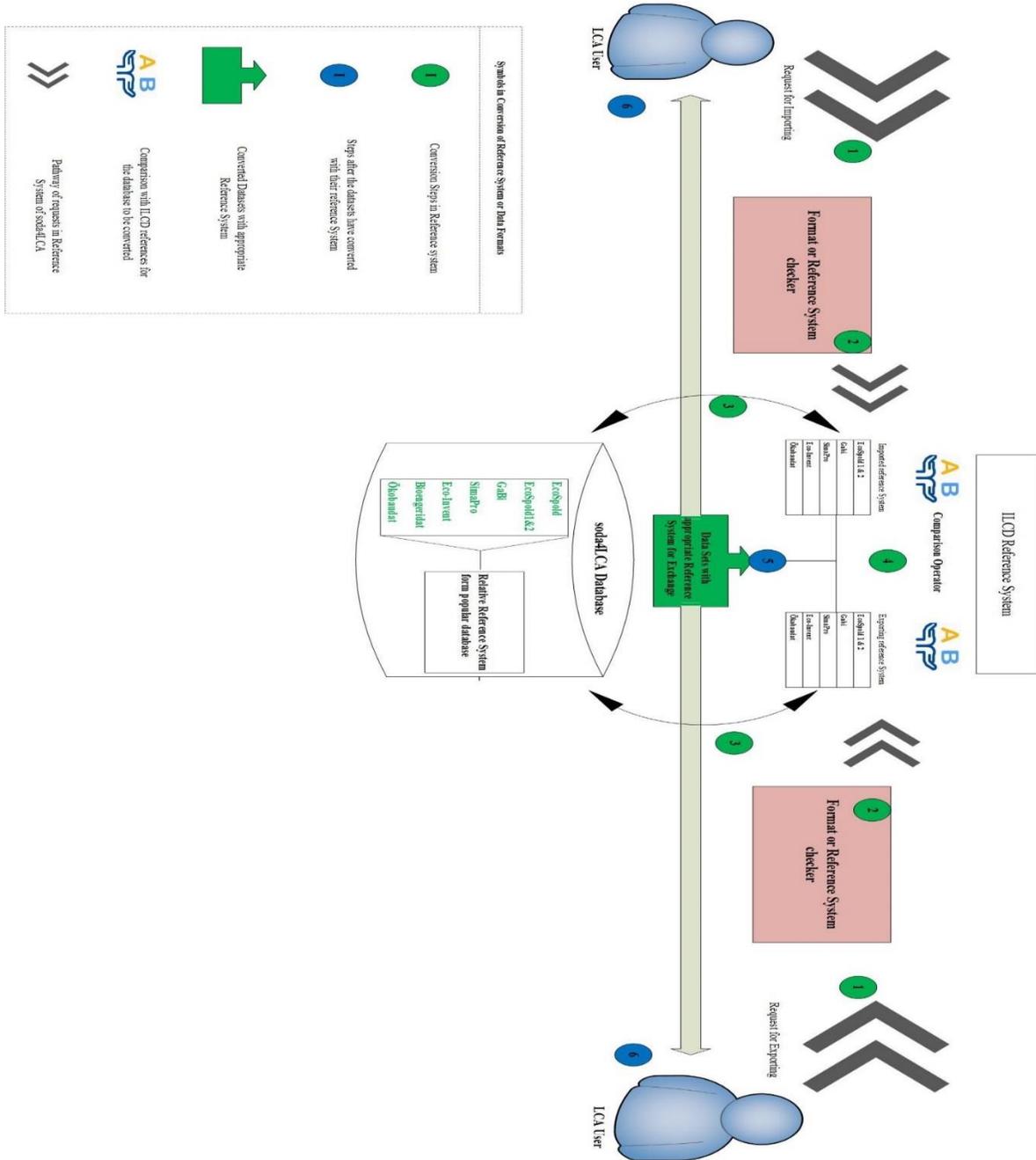
Figure 25: Relationship Model for Conversion of Reference System in soda4lca

the dataset had been fixed, which will be provided to the user. If the format for selected dataset for exchange doesn't exist in soda4lca database, it has to convert the dataset into ILCD reference system which is default and provide the results to the user.

Step4: Conversion of the reference system between different database

This is the crucial part in the reference system, which actually compares the receiving datasets against the ILCD reference system in the first place. It checks for existence of all the references elements such as links, categorization of compartments, flows into appropriate sectors and location based references. If there any missing reference when they have been going to be exchanged between different formats and ensures that all the datasets exist in the proper hierarchy of the reference system from the database trying to communicate.

Let's assume the same example in the step 1, user is trying to import form Open LCA database with its appropriate database reference system. The information about the receiving dataset is provided already in the step 3 for both the cases in importing and exporting. Since the reference system is already known and the architecture for the openLCA reference system is taken from the soda4lca database and place in the comparison table whether to check that all the criteria has been matched after the exchange of dataset has been done.

OpenLCA reference is converted to ILCD reference system by allocating all the appropriate reference and related section based on categorization of ILCD reference system. During this conversion all the missing links to the related datasets have been fixed during the conversion into ILCD. Conversion may take the following steps: [13]

1. Aligning the elements from OpenLCA with ILCD datasets which already exists with a proper unique identifier and nomenclature. Which sorts the most of the difficult issues in conversion process.
2. **Mapping the reference list elements with ILCD reference system.** This is task is more tedious in which all the reference elements between two different formats do not match perfectly and hence there is a need to implement the Mapping table between the formats. The mapping table allows you to allocate the missing reference elements when they have

57

been communicated between inter-formats. Mapping table consists of additional UID's and there reference to the linking datasets. There are major reference elements that need to be taken care such as Unit groups, Flow properties, Flows, Process, Product systems, calculating the inventory, Projects, LCIA methods and Sources.

There are few examples of mapping table for the major reference elements for the conversion from Eco-Spold to ILCD reference system, the information is gathered based on the deep research of OpenLCA format converted, green delta Association.

Mapping of compartments between the databases references, the UID of the Eco Spold is given with a new number for identification along with its version number. Thus the UID has been change for the Eco-Spold datasets when it has been used in ILCD reference system as explained in the figure 24. There are number of compartments and related UID's for the compartments and sub-compartments, but in the figure 24 we are considering only first 10 element to demonstrate the conversion of reference system between different databases for analysis.

1	9ee6ba06-4401-409c-ac4e-e8ec188aa512	2	1	1	air	high population density
2	6a098164-9f04-4f65-8104-ffab7f2677f3	2.1	2	2	air	low population density
3	30347aef-a90b-46ba-8746-b53741aa779d	2.2	3	3	air	low population density, long-term
4	45bb416c-a63b-429f-8754-b3f76a069c43	2.3	4	4	air	lower stratosphere + upper troposphere
5	7d704b6f-d455-4f41-9c28-50b4f372f315	3	5	5	air	unspecified
6	2d0acbd3-2083-4011-9a29-20c626b23dc3	2.4	6	6	resource	biotic
7	ad5c1ad3-29c6-413e-a498-dcfc061f314c	1.1	7	7	resource	in air
8	e47f0a6c-3be8-4027-9eee-de251784f708	1.1.3	8	8	resource	in ground
9	963f8022-3e2e-4be9-ad4d-b3b7a2282099	1.1.1	9	9	resource	in water
10	a119c440-7e83-4655-a874-97fe1468315a	1.1.1	10	10	resource	land

Figure 26: Mapping table for Compartment UID's and Naming conversion [Eco-Spold to ILCD [15]

Mapping of flows between the databases references, there are literally thousands of elementary flows. Thus we demonstrate only few flows to show the way they have been converted by assigning a new identifier for the flow. Figure 25 shows the flows described in Eco-Spold and at the same time figure 26 shows the means and main parameters that are used while converting the reference system from Eco-Spold to ILCD [6].

58

Elementary flows in Eco-Spold are described with the name, UID and their chemical composition as shows in the figure 25, which do not exists in ILCD format because the way expressed in ILCD is totally different than Eco-Spold.

2	Acetaldehyde	000075-07-0	CH3CHO
3	Acetaldehyde	000075-07-0	CH3CHO
4	Acetaldehyde	000075-07-0	CH3CHO
5	Acetaldehyde	000075-07-0	CH3CHO
6	Acetaldehyde	000075-07-0	CH3CHO
8	Acetic acid	000064-19-7	CH3COOH

Figure 27: Elementary flows in Eco-Spold [5]

The converted Mapping table figure 26 shows the conversion of the ILCD UID for elementary flows which are associated with Eco-Spold.

2	fe0acd60-3ddc-11dd-a7a8-0050c2490048	false	1.00
3	e2fa6879-6555-11dd-ad8b-0800200c9a66	false	1.00
4	fe0acd60-3ddc-11dd-a7a7-0050c2490048	false	1.00
5	fe0acd60-3ddc-11dd-a7a5-0050c2490048	false	1.00
6	fe0acd60-3ddc-11dd-a7a6-0050c2490048	false	1.00
8	08a91e70-3ddc-11dd-92c6-0050c2490048	false	1.00
9	d86c88f0-6555-11dd-ad8b-0800200c9a66	false	1.00
824	08a91e70-3ddc-11dd-96e8-0050c2490048	true	1.00
825	08a91e70-3ddc-11dd-96e9-0050c2490048	true	1.00
826	08a91e70-3ddc-11dd-96e7-0050c2490048	true	1.00
827	08a91e70-3ddc-11dd-96e5-0050c2490048	true	1.00
828	08a91e70-3ddc-11dd-96e6-0050c2490048	true	1.00

Figure 28: Elementary flows converted from Eco-Spold to ILCD

Numerical number is the map in the Eco-Spold reference system and stream of characters number shows the new Unique ID's provided in ILCD reference system, the value TRUE or FALSE will be assigned to each Elementary flows showing that True is the dataset to be used in the process documentation and False is a waste that can be a type of flow into any type of different compartments. The Numerical value 1.00 at the end shows the conversion standard for the unit of

measurement related to respective elementary flows, for example if the Acetic acid in Eco-Spold is measured in gram and the same should be done in ILCD too when the reference system has been changed and if it with another numerical value, the same amount of value must be Multiplied.

Mapping of names between the databases references, names used in Eco-Spold and ILCD resembles almost same in format, but in some case Eco-Spold used short naming conventions rather in which ILCD explain the complete name without any abbreviation as shown in the figure 27.

Eco-Spold	ILCD
1,4-Butanediol	butylene glycol
2,4-D	(2,4-dichlorophenoxy)acetic acid
2-Methyl pentane	3-methylpentane
2-Methyl-1-propanol	isobutanol
2-Methyl-2-butene	2-methyl-2-butene
2-Propanol	isopropanol
3-Methyl-1-butanol	2-methyl-1-butanol
4-Methyl-2-pentanol	pentanol
4-Methyl-2-pentanone	methyl isobutyl ketone
Acenaphthene	acenaphthene

Figure 29: Naming conversion between ILCD and Eco-Spold [5]

There are lot of names for the different categories of datasets in each and every reference elements which are expressed in different ways. The names which resemble same will aligned to the same structure for the communicating reference system but some nomenclature convention may differ, thus they have to align using the Mapping table shown in figure 27.

Mapping of unit of measurements between the databases references, figure 28 shows the clear naming conventions and respective maps for the units used in the respective database reference system. Eco-Spold and ILCD both provide different UID's for the unit conventions.

Eco-Spold [2] Unit Convention   ILCD Unit Convention

01092e4c-ce6c-4026-8f85-c1d9da5cdb24	vehicle-km	5454b231-270e-45e6-89b2-7f4f3e482245	Units of items*length	Items*k
0e6f38bc-51be-4382-802d-1f389a88e589	person-km	5beb6eed-33a9-47b8-9ece-1dfe8f679159	Units of items	Item(s)
1017f68a-f818-4f9d-a34d-a31e7386f628	m2	838aaa21-0117-11db-92e3-0800200c9a66	Units of mass*length	t*km
22dd5d20-97ab-471f-aa89-fdb38e46f802	%	838aaa22-0117-11db-92e3-0800200c9a66	Units of length	m
25371fa0-c015-46a9-a4ea-af7e480736f0	EUR2009	93a60a57-a3c3-11da-a746-0800200c9a66	Units of energy	MJ
2cf92850-0f92-4004-9dba-a6ceb6a414c2	year	93a60a57-a3c3-12da-a746-0800200c9a66	Units of volume	m3
2f4daad7-5331-4f14-930c-d8bca924557d	metric ton*km	93a60a57-a3c3-16da-a746-0800200c9a66	Units of radioactivity	kBq
3ce424af-a3eb-4c16-9bc8-aa4443c94f50	kg/s	93a60a57-a3c3-18da-a746-0800200c9a66	Units of area	m2
481b9712-c417-44f1-bfba-38d58088173c	m3*year	93a60a57-a3c3-20da-a746-0800200c9a66	Units of area*time	m2*a
487df68b-4994-4027-8fdc-a4dc298257b7	kg	93a60a57-a3c3-23da-a746-0800200c9a66	Units of volume*time	m3*a

Figure 30: Unit Comparison between Eco-Spold [2] and ILCD [5]

In Eco-Spold, they do not provide the category or the factors while naming the unit of measurements. ILCD specifies the category for the unit belonging and the prescribed unit to be measured for the datasets used in a process.

Eco-Spold UID	ILCD UID	Conversion Value
4923348e-591b-4772-b224-d19df86f04c9	93a60a56-a3c8-17da-a746-0800200c9a66	1.0
5b972631-34e3-4db7-a615-f6931770a0cb	01846770-4cfe-4a25-8ad9-919d8d378345	1.0
77ae64fa-7e74-4252-9c3b-889c1cd20bfc	93a60a56-a3c8-11da-a746-0800200c9a66	3.6
86b0e4a2-57e5-48b8-a43d-22c7f8490dca	93a60a56-a3c8-22da-a746-0800200c9a66	0.0010
86bbe475-8a8f-44d8-914c-e398787e7121	93a60a56-a3c8-19da-a746-0800200c9a66	10000.0
980b811e-3905-4797-82a5-173f5568bc7e	93a60a56-a3c8-11da-a746-0800200c9a66	1.0
ae252091-811b-461b-8e89-8f3075639eb1	838aaa23-0117-11db-92e3-0800200c9a66	1000.0
de5b3c87-0e35-4fb0-9765-4f3ba34c99e5	93a60a56-a3c8-22da-a746-0800200c9a66	1.0
e32b56ef-fa80-4487-9796-f3c1476c27b3	c0447923-0e60-4b3c-97c2-a86dddd9eea5	0.041667

Figure 31: Conversion of Units [5]

During mapping between the two reference system of databases, the UID's form different databases are aligned together with the conversion constant, which shows that dataset with a particular unit is converted from one unit to another there a value change of the particular number called as "*Conversion constant*". Example: when the value for unit Kwh [Kilowatt-hour] in Eco-Spold being converted into ILCD based units of measurements has to be multiplied by 3.6 conversion constant to be aligned with units reference system in ILCD. This has to be done because energy in ILCD reference system is measured in Joules [5].

3. Aligning new datasets or reference elements into database system. There are some scenarios where some of the flows, units, process, products and other additional datasets may not available in Provider database but exist in the user database trying to use those datasets from provider. Thus there is no existence of such an element by the provider and tries to use it during the process may point to null or do not show any results in calculation.

During the assignment of the datasets from one database to another, all the required datasets have to be checked thoroughly before providing to the User. While adding the new datasets in the database for the importing user, it is compulsory to mention the Name of the process, contacts, formulas used in process parameters, Meta information, related documentation for the dataset, etc.

We also mention in Future Work section of the thesis, saying how the user also can add some additional datasets if it is really needed.

Step5: Converted Datasets in Soda4lca

Once the conversion of references of datasets while communication from one database to another is completed. ***The datasets with their appropriate reference system has stored in the part of soda4lca database*** where, the route has been decided to where to migrate the data or for the appropriate user to be communicated.

Since, the converted datasets with their appropriate reference system has been stored in soda4lca. Whenever any other user tries to request the same datasets with a request for similar reference system, then the datasets can be routed directly without any need of conversion once again.

Step6: User Application

The converted datasets will be well developed according to the user requirements. He will be able to work on the same reference system which he was working before with some included datasets having proper reference system. This facilitates the user for the task that he is trying to perform through any LCA node ILCD network of databases.

The tasks that are done by the user can be importing the datasets from a database server or exporting [adding the datasets of LCA] to the database server through soda4lca has to undergo through this Reference system [conversion or alignment of correct datasets] application inside soda4lca database instance.

The above steps clearly describes the conversion that has been achieved between two different formats in reference system that has been implemented in the soda4lca database instance. This technique has been implemented to solve the data harmonization problem and the loss or repetition of data. During the steps of implementation for the better understanding and explanation only required datasets are considered, since it is impossible to show the conversion of all the datasets.

Future Work:

Reference System in soad4LCA is currently implemented for conversion of database references from Eco-Spold 1 and 2 to ILCD reference system and vice versa. This implementation has been explained clearly in the thesis.

Implementation in Other Database References

There are different number of databases Gabi, CSV (SimaPro), DataSmart, PSILCA, Eco-Invent and so on. In soad4lca reference System, architecture of other reference system as mentioned above database is currently not available. But the technique for conversion of the datasets from these different databases also carries the same implementation structure but there may be a variation in the Meta information related to the datasets, referencing ways between their elementary flows and other different criteria as mention in the Thesis work.

Other formats also helps in improving the quality of soda4LCA reference system application, since soda4lca is a database instance which connects n number of nodes in ILCD network and the LCA datasets can be communicated without any data harmonization problem. Currently it is a complex job to store all the Reference System Architecture from all the databases in soda4lca, which is to be continued in future research for developing the soda4lca database instance.

Implementation of Graphical User Interface [GUI] for LCA Users

Soda4LCA has web interface for the LCA users to access the difference datasets all across the Life Cycle Data Network [LCDN]. Due to the current existence of "Reference System Conversion" application in soda4lca as explained in thesis document, it is able to add some additional feathers on the soda4lca web interface.

Where the datasets have been converted between the different LCA formats by LCA users in soda4lca, if the user is not satisfied with current existing datasets there is an option for the user to add a new dataset that he will be using in his process with all the information related with the dataset as shown the figure 30.

As the user enters all the credentials that are required to add a dataset by the application form to add dataset as shown in the figure 30. The credentials are received by soda4lca database, where it is accessible by the admin sector in soda4lca. Admin cross-verifies the requirements of the dataset to be fulfilled and if it does, then it will be updated into the soda4lca database and an acknowledgement is sent to the user. Once the dataset is added in the database, he is able to work with his process and use in any of his LCA calculations.
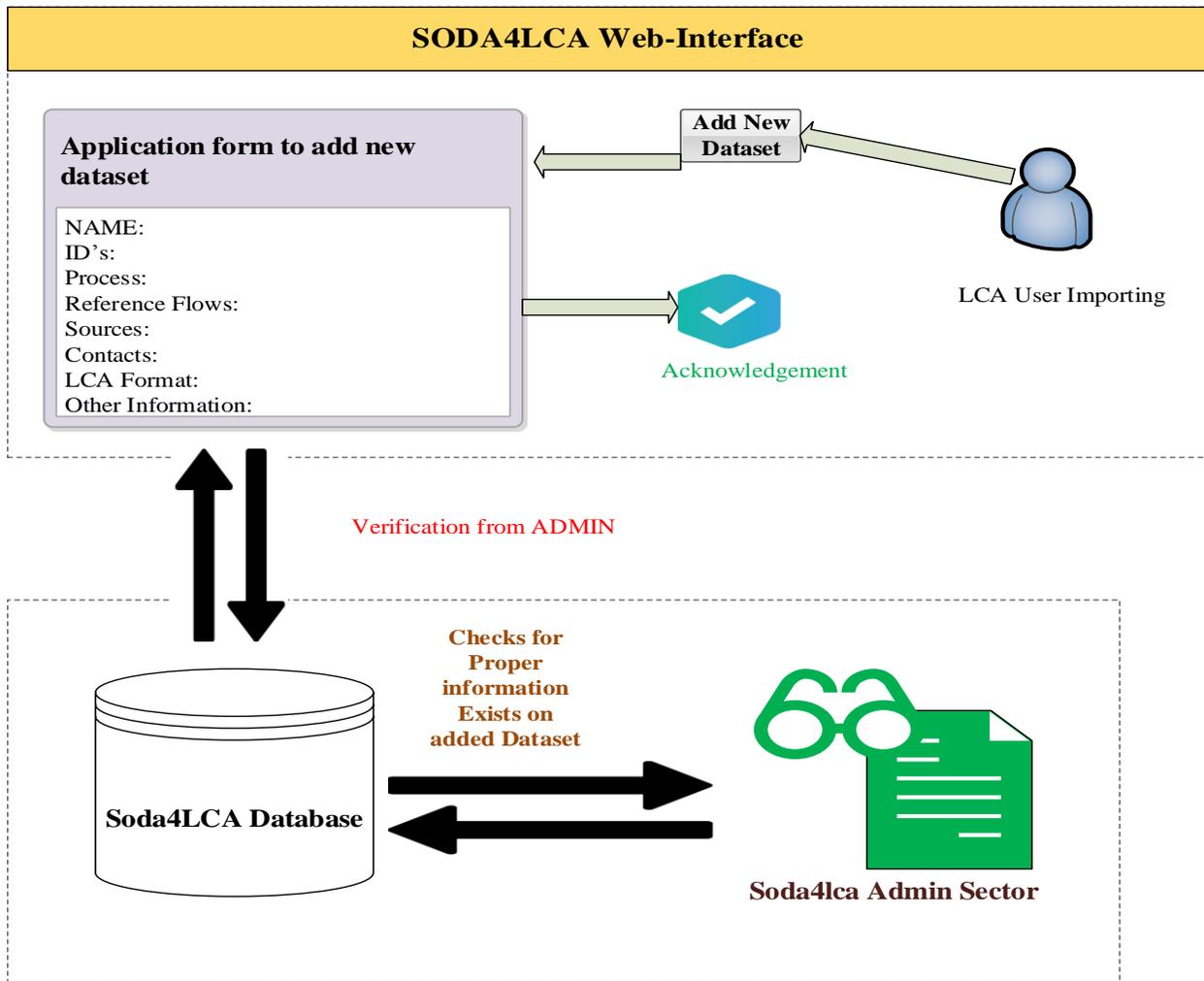
Figure 32: User Option for adding the Dataset

The added dataset and that are uploaded into soda4lca database can be used by any of the other LCA users, who are trying to access the LCDN nodes via soda4lca. This application enables the users to work with their more comfortable and define the information which they are convenient to work with it.

Conclusion

The importance of the work was to increase the efficiency of data exchange between different LCA all across the world. In current situation, it's becoming more difficult to access the LCA data due to existence of more number of databases which uses their own Data Formats or Reference System to organize the datasets in their database. Thus the implementation of reference system supports to solve LCA data harmonization issues and replication of data between different formats, otherwise it would have been enormously time consuming for manually mapping these information during the exchange of data between the users.

There is huge amount of data transaction across the world related to LCA process and manufacturing. These information can be exchanged between the independent users to achieve better performance in their production purposes. LCA resources are available in different databases which are confined to their comfortable data formats or reference systems which are not unique with all the users. The Users wishing to exchange the LCA data between them these inter-format users face a lot of data harmonization issues leading to misplacing of data or providing the wrong information to implement in their process.

Soda4LCA a web-based database instance which connects all these databases which relies on their own data formats or reference system and bringing into on single platform by providing an application called *"Reference System"* to access the LCA back ground data. The implementation of reference system can successfully convert the datasets from one format to another for the user using different databases and their respective data formats. Reference System application will compart any dataset with give its particular format with ILCD Data format as default and makes the necessary changes, finally the datasets are converted into the user required reference system for their database.

Application currently provides the conversion between the Eco-Spold 1 and Eco-Spold 2 to ILCD format and vice versa. The conversion methods for other database reference system carries the same implementation techniques as used to convert the above Data formats, but need to include additional datasets and reference elements which are aligned in their database reference system

according to names and conventions that are used with them. These information have to link with the default ILCD data format used in soda4lca, which helps in further conversion of those data sets with their data formats between other data formats. This is a challenging way to include all the database reference system architectures in soda4lca and increase the data availability between immense numbers of users to be considered as the future work.

To conclude with, the Internet Accessible Reference system to access background LCA data is a reliable application to access huge data repositories for LCA information for all kind of LCA users, those are willing to use different data repositories based on their convenience and need to exchange the data between other database repositories. Application plays a major role in Information Technology sector for solving Data Harmonization and Data Handling issues.

References

[1]   15 March 2016. [Online]. Available: http://www.thelcacentre.com.

[2]   J. E. Commission, ILCD Handbook: General Guide for Life Cycle Assessment: Detailed Guidance, 2010.

[3]   I. 13.020.60 and 13.020.10, "ISO 14040:2006 The principles and framework for life cycle assessment". 01 July 2006.

[4]   J. R. C. -. I. f. E. a. Sustainability, "Compliance rules and entry-level requirements," 2012.

[5]   "Open LCA organization: Datasets and Information," November 2013. [Online]. Available: http://www.openlca.org/.

[6]   D. A. Ciroth, "Working with multiple databases in one LCA software," April 14, 2014.

[7]   "Ecoinvent LCI Database," Ecoinvent Centre, 2011. [Online]. Available: http://www.ecoinvent .org/database.

[8]   W.   link,   "Life-cycle   assessment,"   March   2016.   [Online].   Available: https://en.wikipedia.org/wiki/Life-cycle_assessment.

[9]   KIT, "Soda4LCA Open Source Project: Database Instance for accessing the LCA Data," Infomation of Applied Informatics, Karlsruhe Institute of Technology, Germany, Karlsruhe, 2011.

[10] O. K. Clemens Düpmeier1, "Concept and Implementation of a Service API for the Remote Access to Life Cycle Data Assessment Databases," 2011.

[11] "ILCD Data Network and ELCD Database," JRC European Commission. [Online].

[12] K. I. o. T. G. Institute of Applied informatics, "EPD Data Format – Developer Quick Start Guide:," April, KARLSRUHE, 2016.

[13] C. D. ,. O. K. Marc-Andree Wolf1, "The International Reference Life Cycle Data System (ILCD) Format – Basic Concepts and Implementation of Life Cycle Impact Assessment (LCIA) Method Data Sets," 2011.

[14] C. P. E. K. H.-T. D. B. R. I. Henrik Wenzel-Institute for Product Development, "The Product, Functional Unit and Reference Flows in LCA," 2004.

[15] o. R. C. o. E. Commission, "Datasets from ELCD database," April, 2016.

[16] C. G. O. K. C. S. Clemens Düpmeier, "Managing LCI Data from Different Workgroups within the same Instance of an LCA Database," 2012.

[17] C. Düpmeier and O. Kusche, "Concept and implementation of a service API for the remote access to life cycle data assessment databases.," EnviroInfo 2011 - 25th Internat.Conf.

[18] C. D. a. O. Kusche, "Managing LCI Data from Different Workgroups within the same Instance of an LCA Database," 26th Internat.Conf.on Informatics for Environmental Protection. 2012.

[19] "Mapping Datasets and related information for mapping table in Implementation," 23, October, 2015.