Intensive Data and Knowledge-Driven Approach for Sustainability Analysis: Application to Lignocellulosic Waste Valorization Processes

Jean-Pierre Belaud, Nancy Prioux, Claire Vialle, Patrice Buche, Sébastien Destercke, Abdellatif Barakat, Caroline C. Sablayrolles

Intensive data and knowledge-driven approach for sustainability analysis: Application to lignocellulosic waste valorization processes.

Jean-Pierre BELAUD[1], Nancy PRIOUX[1], Claire VIALLE[2], Patrice BUCHE[3], Sébastien DESTERCKE[4], Abdellatif BARAKAT[3] and Caroline SABLAYROLLES[2].

[1] Laboratoire de Génie Chimique (LGC), Université de Toulouse, CNRS, INPT, Toulouse, France; jeanpierre.belaud@ensiacet.fr, nancy.prioux@ensiacet.fr

[2] Laboratoire de Chimie Agro-industrielle (LCA), Université de Toulouse, INRA, INPT, Toulouse, France; claire.vialle@ensiacet.fr, caroline.sablayrolles@ensiacet.fr

[3] LIRMM GraphIK, INRA, UMR Ingn Agropolymeres & Technol Emergentes 1208, Montpellier, France; patrice.buche@supagro.inra.fr, barakat@supagro.inra.fr

[4] CNRS UMR Heudyasic, rue Personne de Roberval, F-60200 Compiègne, France; sebastien.destercke@hds.utc.fr

Corresponding author: jeanpierre.bealud@ensiacet.fr

Abstract

The use of circular economy is becoming more and more important, particularly in the field of agriculture, a major provider of waste. In particular, a lot of researches are being done to transform the lignocellulosic waste from agriculture through desired "sustainable" processes. Sustainable processes mean economically viable, socially accepted, and environmentally responsible processes. Thanks to the "life cycle thinking", it is possible to assess such potential environmental impacts. However, these environmental analyzes require a lot of specific data, whose collection can be long and tedious, or simply impossible in practice. On the other hand, the huge amount of scientific articles describing the processes of valorization of co-products of agriculture constitutes a great, largely under-exploited source of data. Knowledge engineering (KE) tools can be used to compile processes and analyze them. In this paper, we propose an innovative approach, based on intensive data and KE methods, to help a decision maker to choose between different pretreatment processes and different biomasses. The main goal is to develop an intensive, semi-automated data collection approach and an associated tool for assistance with choices in a circular economy context. It is defined by five steps: (1) goal and scope, (2) intensive data and knowledge structuration and integration, (3) life cycle inventory (LCI), (4) sustainability assessment and (5) analysis and ranking. The study of 13 pretreatment processes of rice straw and corn stover validate our proposal.

Keywords: Agricultural waste, Circular Economy, Knowledge Engineering, Big data, Life Cycle Assessment, Lignocellulosic biomass

Statement of Novelty

Sustainability analysis need a lot of data. However, the structuration and integration are complicated. A new intensive data and knowledge-driven approach for sustainability analysis is presented.

1 Introduction

The expectation of inhabitants cannot allow keeping the linear "take, make and dispose" pattern. This is why the European Commission has proposed in 2015 the use of the circular economy model to boost the use of sustainable models. The circular economy concept is defined as "one that is restorative and regenerative by design and aims to keep products, components, and materials at their highest utility and value at all times, distinguishing between technical and biological cycles" [1]. This new business model for more sustainable development helps to reconcile environmental, economic and social aspects. The origins, the principles and the limitations of circular economy (CE) models are discussed through few articles and transcribed by Ghisellini et al. [2]. In 2015, the French government proposed the SNTEDD (National Ecological Transition Strategy for Sustainable Development) which is consisted of nine areas, one of which is CE. According to the French Environment and Energy Management Agency (ADEME), CE takes into account three action fields: (1) consumption through the demand and consumer behavior, (2) supply and economic players and (3) waste management [3]. These three action fields account for the entire life cycle of a product, a service or a process. To obtain sustainable models, life cycle thinking can help improve environmental performance and optimize the economic and social benefits. One particular domain where CE and life cycle thinking grew these last decades is the agriculture.

Globally, the population generate 2 000 million ton of agricultural waste per year [4]. The increase of waste production will accompany the projected increase in the world's population. Moreover, human activities decrease the amount of land available for agriculture, which inevitably has impacts on agricultural systems. For Garnett et al., the best approach for the future of agriculture are new agricultural technologies which will facilitate sustainable intensification [5]. Nevertheless, this intensification will lead to more waste of products and resources [6]. According to Horton et al. [7], a major challenge in attempts to achieve sustainability is the parametrization of waste in agriculture. Two classes of waste can be identified: waste from inputs, such as fertilizer or water, and process waste. The process waste comes from the biomass incomplete conversion or material transformation in the supply chain that goes from agricultural production to food consumption and is mostly composed of lignocellulosic by-products. Lignocellulosic biomass is one of the most abundant and cheapest renewable resources on Earth. The production of biomaterials, biomolecules and bioenergy is based on the lignocellulosic biomass bioconversion, which involves enzymatic hydrolysis of the biomass to release glucose. The lignocellulosic biomass is composed of four main components: lignin, cellulose, hemicellulose, and phenolic acids but only cellulose and hemicellulose, can be hydrolyzed to generate glucose. Although lignocellulosic biomass is a renewable resource, the processes for transforming this biomass must be sustainable to participate in overall sustainability. That is why more and more agri-food processes integrate sustainability assessments [8, 9]. To generate good glucose yields, it is essential to plan pretreatments of lignocellulosic biomass before its enzymatic hydrolysis. In the 30 recent years, numerous pretreatment processes have been studied and published [10]. Various factors have been used to compare the performance, efficiency or environmental impacts of the pretreatment [11]. The environmental factors, energy consumption and energy efficiency may be considered to be classical factors [12–14]. However, criteria are lacking to guide the choice between all these processes. Using environmental, economic and social assessment in a CE context is a good way to guide the choice. In this paper, the social dimension is not under consideration. Economic and environmental assessments in a CE context rely on methods such as life cycle assessment (LCA) and life cycle costing (LCC). These assessments need many data particularly process data that may be found in the scientific articles. In order to feed our lignocellulosic waste valorization application, the articles describe pretreatment processes. However, there is a lack of methods for the use of these semi-structured data for completion of the life cycle inventory (LCI). Use of methods coming from intensive data and knowledge domains allows the development of an "augmented approach" for economic and environmental impact analysis. **Figure 1** shows a synthesis of our proposal.

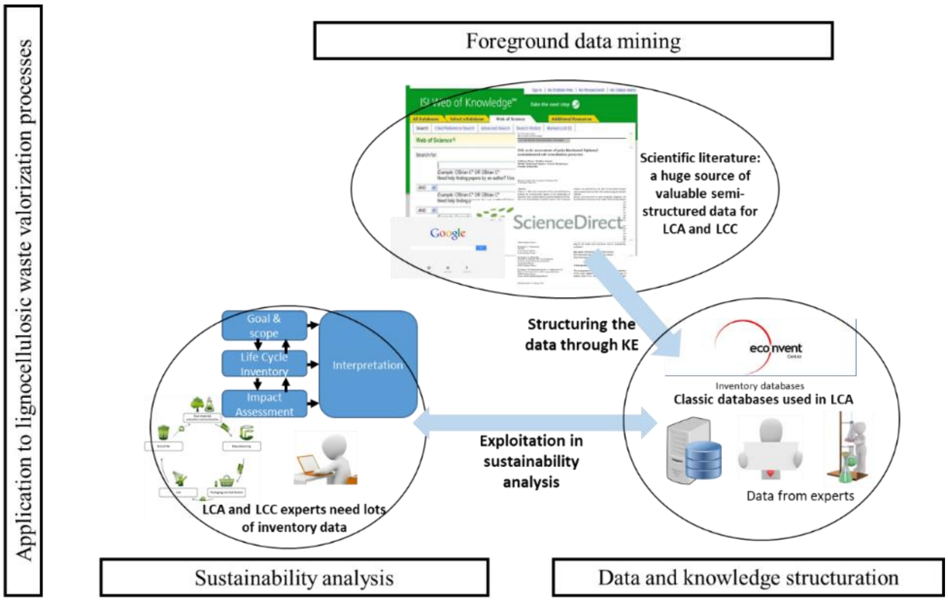


Figure 1: *Intensive data and knowledge-driven approach for sustainability analysis*

This paper presents an approach that helps to analyze different pretreatment processes technological paths and different biomasses in a CE context. Our approach for sustainable analysis is driven by intensive data and knowledge. Indeed, sustainability analysis (here LCA and LCC are selected) needs for process data and cost data. The scientific literature can provide these data and knowledge engineering methods (KE) can bring the structuration of data and knowledge. Many methods from big data allow structuring data and knowledge. In this paper, KE is used. KE structures knowledge into formal representation for computing thanks to a standard vocabulary. The goal of this method is (1) to place LCA upstream in eco-design processes and to support the selection of a unitary operations chain and biomass and (2) to make use of intensive data for given process resulting from experiments performed by researchers around the world. An additional benefit of this method is to avoid the need to perform some time-consuming and expensive experiments. The second aim in this paper is to demonstrate the feasibility of a pipeline (detailed in **Section 3**). Process data as inputs are found in scientific documents and final output is a ranking of those processes based on sustainability indicators.

After a discussion on LCA, LCC and KE methods and their coupling in the literature (**Section 2**), our approach for sustainability analysis is spelled out (**Section 3**). The approach is deployed with an agricultural wastes valorization (**Section 4**): environmental analysis for six pretreatment processes and two lignocellulosic biomasses (rice straw and corn stover). This paper finishes with conclusions on our developed approach and gives perspectives.

2 Methods and tools

In our approach, we select the life cycle assessment method which is an ISO method [15]. LCA evaluates the potential environmental impact of a product or service over its entire life cycle [16]. The life cycle of a product/service can be broken down into several steps, beginning with product design and ending with waste disposal or product recycling, after various stages of transformation and use. The life cycle assessment method includes a number of flows, which can be classified into two groups: (i) elementary flows and (ii) intermediate flows. Elementary flows involve exchanges with the ecosphere: the extraction of the raw materials (gas, minerals, etc.) and the emission of pollutants. Intermediate flows are the flows of energy or matter between steps. The first stage of LCA is the definition of goal and scope. This stage is very important as it identifies the issue considered and defines the boundaries of the system. The functional unit (FU) is defined at this stage. Care should be taken with this definition, as it can influence the results of the LCA [17]. The second stage of LCA is the establishment of a life cycle inventory (LCI). The LCI is a listing of the amounts of pollutants emitted and the resources extracted throughout the life cycle of the product or service concerned. This inventory is generally split into two parts: the inventory of the background system and the inventory of the foreground system. The foreground system corresponds to processes under the control of the decision maker, for whom LCA is carried out. The background system consists of all other processes interacting directly with the foreground system [18, 19]. The data for this inventory may be obtained directly, by on-site measurements (primary data), or indirectly, from published scientific articles, models, and databases (secondary data). The inventory of the foreground system is generally based on primary data, whereas that of the background system relies on secondary data sources [20]. When primary data are missing, the typically huge number of articles describing process operations could provide a valuable source of data for the foreground system. Yet, prospecting a huge quantity of unstructured data cannot be done without some degree of automation. So, to use this data in the foreground system, a method must be created. The third stage is life cycle impact assessment (LCIA), in which the numbers of pollutants and resources listed in the LCI are translated into environmental impacts [21]. The last stage is results analysis and interpretation which consists of identification of the significant issues based on the LCI and LCIA results, evaluation of the sensitivity of these issues, checking of consistency and completeness, and conclusions, recommendations and limitations.

c

The main goal of semantic web-based knowledge engineering methods (KE) is to structure the experimental information and express it in a standardized vocabulary. Large amounts of data (as in the "big data" context) expose the limitations of standard statistical software resulting. So data structuring is important [22]. Such structuring can be done using an ontology (the semantic part of our model) to represent the experimental data of interest (**Figure 2**). Ontologies are knowledge representation models that facilitate linkage of open data and offer automated reasoning tools [23]. Once structured in ontologies, collected information and data are made homogeneous and can be processed to the sustainable analysis [24]. This methodology is based on the use of linguistic or syntactic patterns [25] and the extraction of *n*-ary relations. naRyQ (n-ary relations between quantitative experimental data) core ontology has been designed to annotate data tables representing scientific experiment results in a given domain [26]. The core ontology is composed of three kinds of generic concepts: (1) simple concepts, which contain the symbolic concepts (studied objects) and the quantities, (2) unit concepts that contain the units used to characterize the quantities and (3) relations, which allow n-ary relationships to be represented between simple concepts. The core ontology is generic. The concepts belonging to a given domain ontology, called specific concepts, must be defined and appear in the ontology as sub-concepts of the generic concepts. The extraction is divided into three steps:

(i) The identification of entities based on knowledge representations, such as ontologies or dictionaries;
(ii) Identification of the trigger word for the relationship, through the use of dictionary-based methods or rule-based approaches to construct patterns [27], or with machine learning methods [28];
(iii) The construction of binary relationships involving the trigger word and the use of machine learning methods to determine whether the binary relationships concerned belong to the *n*-ary relationship of interest.

In knowledge engineering, the automatic extraction of relevant information from the text and tables of scientific articles is an area of active research. Off-the-peg tools are not yet available, but increasing numbers of ontologies are emerging for the organization and sharing of knowledge in particular domains, and such extraction tasks are performed in various applications. One example, a French ontology, [MS]²O, clusters data relates to transformation processes in food science [29].This ontology allows different teams to work on the same subject and to group their data together in a single database, making it possible to compare different production scenarios. Another example is provided by Rosanne developed as an Excel "plug-in" and constructed from an ontology of quantities and units of measure [30].

Big data technologies can, and have been applied to industrial ecology. Xu et al. [31] explored the possible contribution of big data to industrial ecology through several examples combining these two domains. Combining the huge amount of data available with KE techniques for their exploitation would clearly be beneficial for LCA, as it would make it possible to obtain surrogate data in situations in which specific data cannot be collected, rather than relying on default values. A couple of studies have already explored such a pathway, but work in this area remains limited. For instance, Cooper et al. [32] used big data to complete the background system database. Big data have also been used for LCA in data-intensive life cycle assessment (DILCA) [33], which makes use of KE-based approaches to adapt LCA to technological developments, which may modify LCA results for a given product over time. Finally, KE can also be used for the benefit of LCA through the use of ontologies to represent the life cycle of a product and its LCA [34]. These ontologies represent all the intermediate flows, emissions and extractions. Hence, the KE can structure the data from heterogeneous sources. No articles presenting KE methods and LCC were found but few articles show the cost of ontology engineering in a project or the costs of process material in an ontology. Zhou et al., for example, created an ontology which takes into account operating costs, labor costs and capital cost [35]. ONTOCOM is a reliable cost estimation method for ontology development projects created by Simperl et al. [36]. Lee et al. proposed an ontology for project planning and notably the cost project [37]. An ontology-based approach supporting holistic structural design with the consideration of safety, environmental impact and cost, created by Zhang et al., is the KE approach which uses the most environmental impacts and life cycle cost [38]. However, LCC is not complete and environmental impacts used did not come from LCA. Hence, data intensive and knowledge methods like KE can facilitate the use, for sustainability analysis, of the huge wealth of data available from scientific publications. The developed approach in this paper is particularly applicable to studies where few or no primary data are accessible for the foreground system, because the product is still in the design phase, or because the primary data are proprietary. We therefore decided to use secondary data and KE methods as a source of information for the foreground system. The term "background data" is used here to designate the data describing the background system, and the term "foreground data" is used to designate the data describing the foreground system mainly the process data (material and energy flows, parameters, technologies, …). The approach aims especially to researchers and development engineers for supporting preliminary decisions with respect to sustainability analysis. Intensive data and knowledge-driven approach for sustainability analysis

3 Intensive data and knowledge-driven approach for sustainability analysis

In this section, the general pipeline is developed, while an application is presented in the next section. It consists of five stages. The processing pipeline presented here is based on the LCA method, combined with intensive data and KE methods to complete the data collection for the foreground system. This data collection is a substep of the life cycle inventory. The resulting pipeline, shown in **Figure 2**, has five main stages:

1. Definition of the goal and scope of the study
2. Data intensive and knowledge structuration and extraction
3. Life cycle inventory
4. Sustainability impacts assessment
5. Analysis and ranking

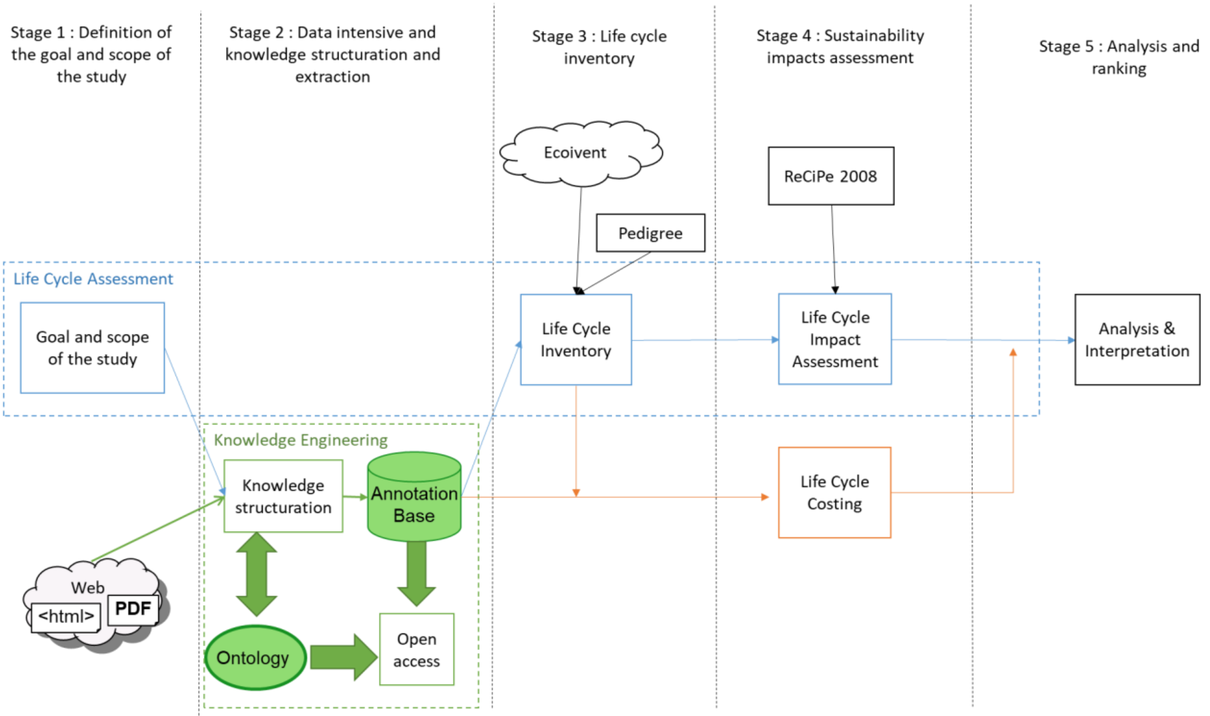


Figure 2. *Pipeline of intensive data and knowledge-driven approach for analysis assessment*

To our knowledge, this is the only pipeline to date to make use of these methods in this way. J. Cooper et al. [32] use KE to complete the LCA, yet with the goal to complete the data collection for the background system and not the foreground system. The order of these five stages must be respected, but it is always possible to return to the previous stage. Indeed, such iterations are even recommended, as they can be used to adjust the data and the methodology, resulting in better results.

3.1 Goal and scope (Stage 1)

The first stage of our method is the goal and scope, which can be split into several different stages. The first substep is the definition of the goal of the study and that the study is being performed for, known as the recipient. In our approach, the recipients are the researchers or research and development engines using the pipeline. This pipeline is particularly suitable for research use because of the system boundaries and the laboratory scale of data collection in published papers. It could easily be scaled up for industry [39]. The second substep is the definition of the boundaries system. It is recommended to use the life cycle thinking (LCT) for the system boundaries. Indeed, the LCT and so the CE encourages a “from cradle to grave” or “from cradle to cradle” approach [40]. However, in the sustainability analyzes it is difficult to integrate downstream elements leading to a preference for “from cradle to gate” approach. These limits must be defined with precision because they are very strong effects on the sustainable assessment. In our case study for example, the inclusion of the upstream biomass supply chain can change the results. The last substep is the definition of the functional unit, the data required, the choice of impact categories, the process tree, with inflows/outflows, and the type of cost sources. The functional unit depends on the goal of the study and the type of process comparison that researcher wants. The choice of impact categories must be justified. The type of cost sources depends on the study, they can be issued from private databases or web public, for example. This first stage also constrains and guides the creation of ontologies in the next stage. Indeed, the definitions provided already structure the knowledge and narrow down the selection of scientific papers and data required for the study, overcoming the need to search the whole worldwide web for data. This stage is done manually by industrial engineers (or process engineers in our example) and sustainable engineers.

3.2 Intensive data and knowledge structuration and integration (Stage 2)

The structuration and integration of the intensive data - and express them in a standardized vocabulary - are done thanks KE methods. This stage, which is divided into several substeps, is derived from KE methods. Heterogeneous experimental data from a vast array of scientific papers are integrated with the @Web. @Web (for Annotated Tables from the web) which relies on an Ontological and Terminological Resource is a collaborative platform to share documents with annotated tables [41]. @Web was developed by the French National Institute for Agricultural Research (INRA). This free, open-source tool can be used for all substeps of data integration of stage 2 of the methodology [42].

The first substep is the selection of documents describing the various processes that we have to compare in the analysis. Experts (particularly process researchers or engineer) identify all published article thanks to different keywords in scientific databases, such as Web of Science or Science Direct. These articles are then sorted by topic, with each topic corresponding to a different type of process that we have to compare in the analysis. An article, which describes different processes, can be sorted in two or more topics. Documents can be uploaded directly into @Web from a desktop or from a collaborative repository management system. Bibliographic references and their entire text, in HTML and PDF formats, are managed by @Web.

The second substep is the assessment of document reliability. A document analyst first enters meta-information to calculate the reliability score: public meta-information, such as the data source (source type, reputation and citation data), and meta-information from Web of Science relating to the data production methods and statistical procedures. @Web proposes a reliability estimation tool [42], with lower scores for the most reliable documents. When knowledge is insufficient (missing information about statistical procedures, publication too recent for a meaningful number of citations), the score is given as an interval between the “worst” and “best” possible reliability of the article, with the width of this interval reflecting the amount of missing meta-information. This reliability score (or range) is completely configurable by the analyst, who is free to change the parameters taken into account in the calculation of the score. These parameters are also completely adaptable: it is possible to add or delete parameters following the study or to modify the influence of one parameter in the global reliability score.

The third substep is the creation of an ontological and terminological resource (OTR) to facilitate the use of data from heterogeneous sources and to guide scientific data annotation. This OTR distinguishes between the concept (it is the conceptual component) and its linguistic expression in different languages (it is the terminological component) [43]. In the conceptual component, the representation of an experiment is given as an *n*-ary relationship between a given result and several experimental parameters. These *n*-ary relations are used to create annotated tables. For example, if the example of a generic *n*-ary relation in the **Figure 3** represents the *n*-ary relation *Unit_operation_relation*, the column of the annoted table which describes this unit operation, correspond to the arguments of the relation *Unit_operation_relation*. The OTR is composed of a core ontology and a domain ontology. The core ontology is composed of the generic concept relation, generic concept dimension, unit concept and quantity concepts. The domain ontology contains specific concepts of a given application domain: all the *n*-ary relation describes the process. Once the OTR has been created, tables of selected documents can be extracted and annotated.

The forth substep is the table extraction then the table annotation. The table extraction corresponds to the extraction of the data tables from HTML version of documents using tag analysis. Then it is the manual semantic annotation of the selected datable using the concepts of the OTR which is done. This step is the annotation table: the annotator selects from the *n*-ary relation concepts defined in the OTR those relevant to annotate table.

The last stage is storage in the annotations base. The annotated data tables are stored in an RDF (Resource Description Framework) triple store, making it possible to use the querying interface (**Figure 3**). RDF is a standard mode of data interchange via the Internet. It is used as an interface between users and the OTR, enabling users to interrogate the OTR in various ways. Users can use a querying tool to rank the data in a specific order, on the basis of data source reliability, for example, or by selecting a kind of process. The annotations base contains the foreground data required for the life cycle inventory generated in stage 3. The document reliability score can be used, in the last stage, to rank the results, to guide the researcher's choice or to delete some non-relevant article.

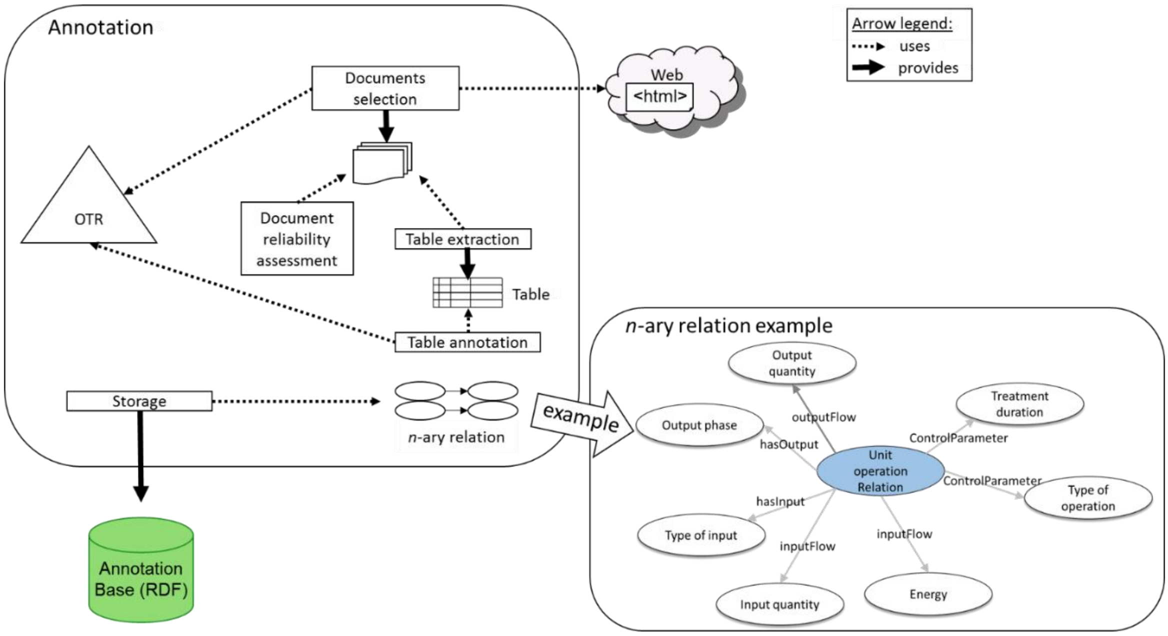


Figure 3 : *The creation of an annotations base in @Web, adapted from with a generic n-ary relation example*

3.3 Life cycle inventory (Stage 3)

The life cycle inventory lists and quantifies the various relevant inputs and outputs. ISO standards [15, 16] describe the different stages of the LCI: data collection, data calculation, and the allocation of flows and releases. Data collection can be split into two parts: data collection for the foreground system and data collection for the background system. Many methods are available for compilation of the LCI and for the organization of these data [21]. Hence, ISO has developed a technical specification for data documentation formats for the LCI [44]. Thanks to this standard, all the LCI background databases, such as EcoInvent [45], the US Life Cycle Inventory Database [46] and the International Reference Life Cycle Database (ILCD) [47] use the same data format. Such formatting is useful, as it is simple to fill out the corresponding form, even if many items are not completed, many chemical species are missing and the intermediate flows of many processes are not available. In this work, EcoInvent was used for background data, and the methodology developed concerns the foreground data.

In this case, foreground data are extracted from the Internet in **Stage 2**. **Figure 4** illustrates the relationships between the different stages and the tools used to perform them. The annotations base (also called RDF) provides foreground data related to the unit process and the functional unit. The background data are obtained from the EcoInvent v3 database. Data are validated by a specific data validation method recommended by ISO 14044 [15], the pedigree matrix approach.

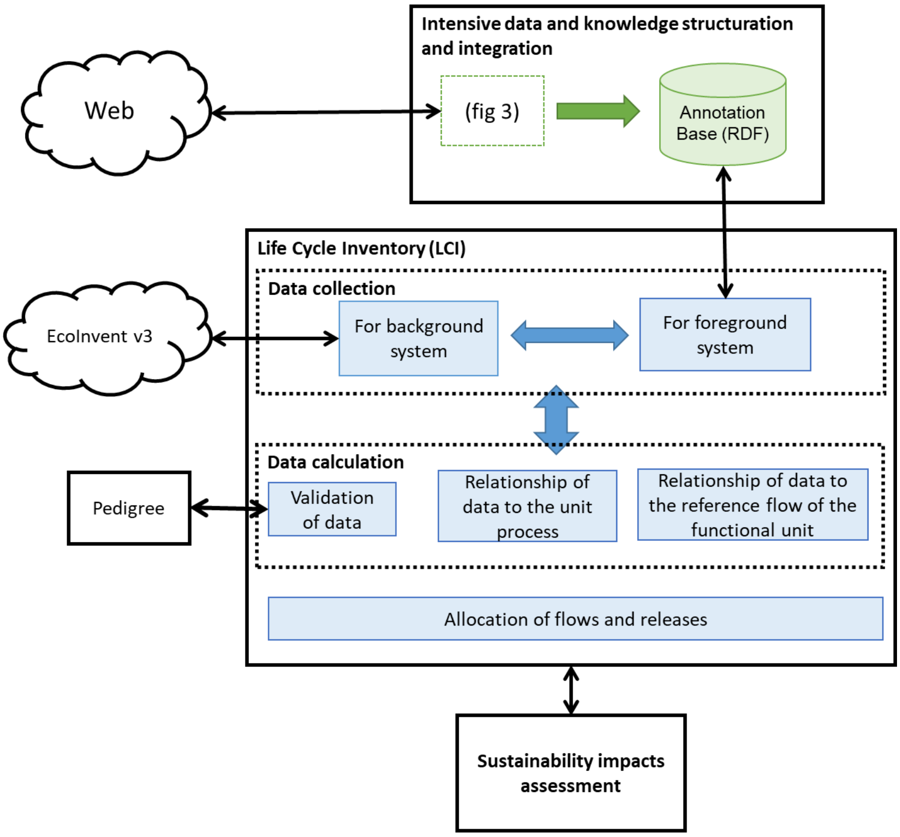


Figure 4. *Stages of the life cycle inventory and connections with preceding and subsequent stages in the pipeline*

We expect the result of this stage to be greatly improved by our proposed method. This new method should increase the completeness and reliability of existing data, by seeking external data from other scientific publications, and should also provide researchers with results without the need to perform another experiment, even in the absence of data. In such situations, it is crucial to evaluate the quality of the collected data. We therefore think that the validation tools provided by @Web are extremely useful, even when not strictly necessary. Evaluations of reliability or relevance may help the analyst to scan collected papers by ranking them, making it easier to retain only the most relevant and reliable data. In some cases, the results may also lead to the analyst returning to previous stages, to redefine the LCA in light of new information.

3.4 Sustainability impacts assessment (Stage 4)

This stage is divided into two parts: the LCA and the LCC. These two assessments can be done separately or combined. Life cycle impact assessment aims to transform inventory results into environmental indicators (also known as impact categories). The set of indicators proposed by the LCA methodology is distributed over three levels of aggregation. The first level concerns the quantifiable physical, chemical and biological effects of the flows of material and energy between the system studied and its environment. The corresponding indicators are called midpoint categories. These midpoint categories depend on the evaluation method used. LCI results are initially classified by midpoint category. The results are then multiplied by impact factors to obtain midpoint scores. The impact factors are derived from various midpoint methods, such as ReCiPe 2008 Midpoint [48] and ILCD 2011 Midpoint [49]. The second level concerns the damage to several areas of protection caused by these effects. Many methods, including ReCiPe 2008 and ILCD 2011, consider three areas of protection: human health, ecosystem quality and resource depletion. The intermediate damage factors are derived from the different methods. However, some methods, such as the ReCiPe method, do not have these factors and use the LCI results directly to calculate the damage score, which is also known as an endpoint score. A single score (third level of aggregation) is obtained by normalizing damage scores with normalization factors. These normalization or weighting factors are provided by the different methods or can be calculated. This stage is generally performed with dedicated LCA software, such as SimaPro® [50] or Gabi® [51].

LCC aims to calculate the sum of the costs during the life cycle. The different costs taken account into this LCC depend on the first stage (**Section 3.1**), especially the limit of the system. Indeed, following the limit of the system, it is possible to take into account the infrastructure costs, disposal costs or biomass costs, for example. Costs like manufacturing costs or initial costs (equipment investment costs) must always be in the LCC even if the systems limits can change them. Indeed, some equipment can be already existing in the infrastructure or it is a new infrastructure all the equipment must be buying. We do not propose a specific method to calculate LCC indicators in this paper because methods depend on the goal and the limit of the study done.

3.5 Analysis and ranking (Stage 5)

Finally, the analysis and interpretation of results are done. This stage is highly linked to the user and the goal of the study and provides decision support. For the LCA indicators, the user can choose between the first, second and third levels of aggregation, according to the type of decision to be taken. The LCC indicators chosen can be coupled with LCA indicators or use separately. The ranking obtained should reflect both uncertainties in the collected data and the reliability of data (using results from @Web), and should be easy for users (in our case, mostly researchers) to read, to facilitate decision-making. We expect the additional information provided by the KE approach to be instrumental in the completion of this task. This stage can be used to different types of LCIA visualizations with news information following the goal of the study. The visualization should never be ignored [32]. Decisions may be taken manually or with the assistance of decision support tool using multi-criteria models, such as sorting models (ELECTRE, PROMETHEE) or aggregation tools [52]

4 Lignocellulosic waste pretreatment processes for biomass valorization

The pipeline presented above is illustrated by the example of lignocellulosic waste pretreatment. The pipeline was developed for the comparison of different pretreatments of two biomasses and different biomasses subjected to the same pretreatment. Below, we focus on rice straw and corn stover pretreatments. Only environmental assessment (LCA) is detailed here. Lignocellulosic biomass, the essential component of woody cell walls in plants, is one of the most abundant and cheapest renewable resources on Earth. The bioconversion of lignocellulosic biomass is a promising method for the production of bioenergy, biomolecules or biomaterials. This bioconversion involves the enzymatic hydrolysis of the biomass to release glucose. The lignocellulosic biomass has four main components (e.g., cellulose, hemicellulose, lignin, and phenolic acids), of which only two, cellulose and hemicellulose, can be hydrolyzed to generate glucose. The lignin and phenolic acids are thus responsible for the recalcitrance of cellulosic materials, the crystallinity of cellulose and the particular surface and porosity characteristics of matrix polymers. Biomass pretreatment is, thus, essential, to decrease crystallinity, to increase the specific surface area and porosity, and to separate out the major constituents. It exists in different types of pretreatment processes that all have their own particularity. **Figure 5** presents an example of process.

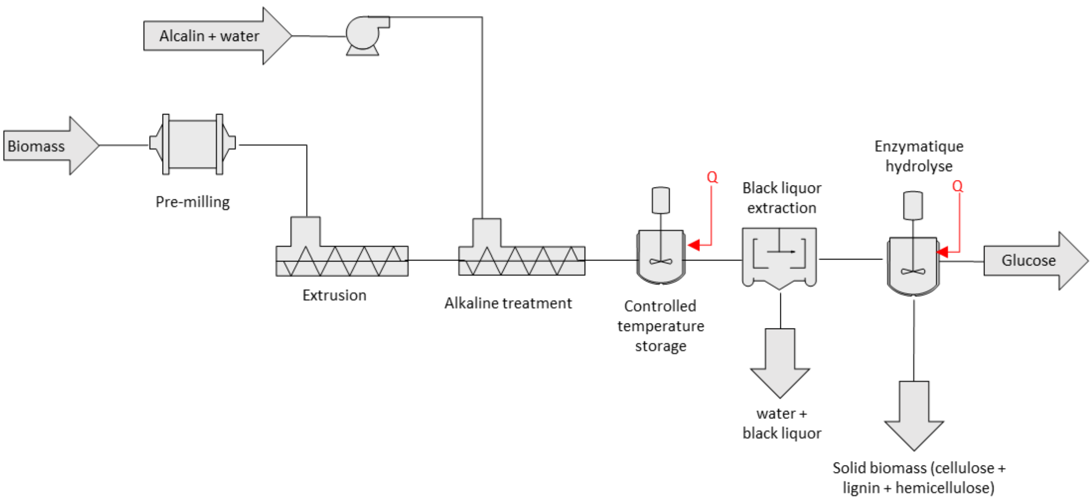


Figure 5: *Example of pretreatment process adapted for Liu et al. [53]*

4.1 Goal and scope

The goal of the study is the comparison between corn stover and rice straw pretreatment for glucose production. Here, the recipients are the researchers. The process of biomass pretreatment is a cradle-to-gate approach [40], extending from the milling of the biomass to its enzymatic hydrolysis (**Figure 5**). Input production (including biomass) is included, whereas the transport of inputs and outputs is excluded. Energy input is a key parameter of the LCA, but data for this parameter are almost always missing. The reasons for this may include the laboratory scale of the selected studies and the type of article where the process is described in which the energy input of a process is rarely measured. Efforts have been made to compensate for this problem by estimating milling energy by the application of statistical models based on classical regression to complementary data extracted from published studies. This regression was based on complementary experiments performed by A. Barakat [41]. Solid and liquid outputs are considered to have no impact, because they contain molecules that may be valorized, and are, therefore, not wasted. The function of the system is glucose production and the functional unit is the “production of 1 kg of glucose.” All results are expressed in terms of this functional unit. Different kinds of processes were selected: for the rice straw the process named PRS and for the corn stover PCS.

4.2 Data and knowledge structuration

The first substep is the selection of documents describing the pretreatment processes of the rice straw. Biomass pretreatment experts identify all published articles corresponding to keywords: “rice straw”, “corn stover”, “treatment”, “hydrolyze” and “milling” in scientific databases, such as Web of Science or Science Direct. They identified 20 relevant scientific articles. These articles are then sorted by topic, with each topic corresponding to a different type of process (**Table 1**). Six types of pretreatment are described in these articles: pre-milling pretreatment (PM), pre-milling and ultra-fine pretreatment (PM-UFM), pre-milling, physicochemical and press separation pretreatment (PM-PC-PS), pre-milling, physicochemical, ultrafine milling and press separation pretreatment (PM-PC-UFM-PS), pre-milling, physicochemical, extrusion and press separation pretreatment (PM-PC-EX-PS) and pre-milling and ultrafine milling pretreatment (PM-UFM). These documents are uploaded in @Web and their meta-information are manually entered. The second substep is the assessment of document reliability. The reliability score of each document is assessed and the results are visible on @Web. All the views from @Web presented here relate to a paper on physicochemical and press separation pretreatment (PM-PC-PS) [42].

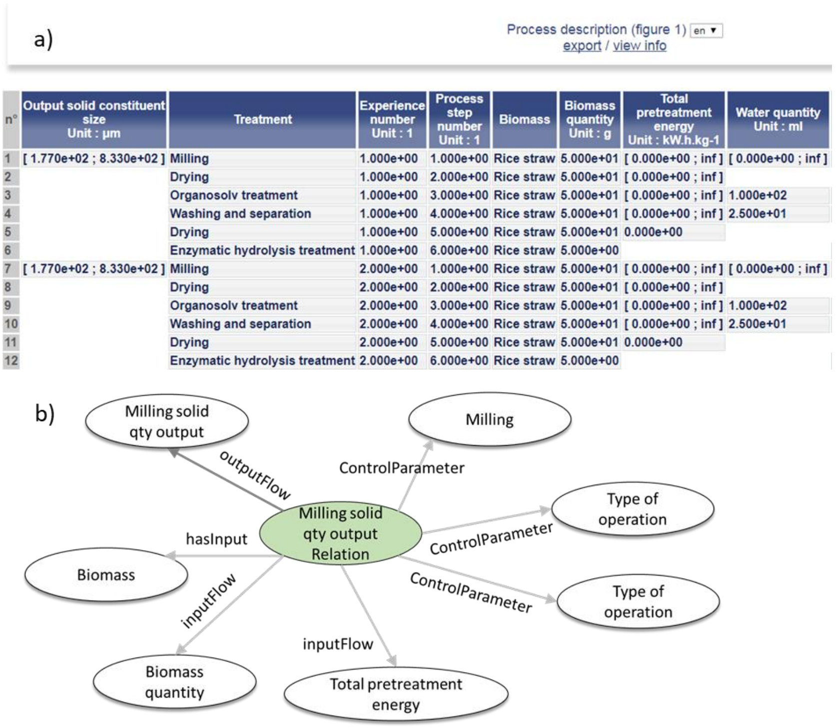
Abbreviation	**Complete noun**
PM	pre-milling pretreatment
PM-UFM	pre-milling and ultra-fine pretreatment
PM-PC-PS	pre-milling, physicochemical and press separation pretreatment
PM-PC-UFM-PS	pre-milling, physicochemical, ultrafine milling and press separation pretreatment
PM-PC-EX-PS	pre-milling, physicochemical, extrusion and press separation pretreatment
PM-UFM	pre-milling and ultrafine milling pretreatment

Table 1: *Summary of studied pretreatment types*

The next step is the creation of OTR. The conceptual component of Biorefinery OTR is composed of a **core ontology** to represent n-ary relations between experimental data and a **domain ontology** to represent specific concepts of a given application domain – here the biorefinery**.** In Biorefinery OTR, every relations represent either experiments that characterize biomass or experiments involving unit operations performed on biomass. An example of an n-ary relationship is provided in **Figure 6b**. It represents the milling experimental result for a given biomass. It is characterized by 7 arguments of which is the given biomass quantity and another the milling solid quantity output.

The forth substep is the table extraction then the table annotation. The table extraction corresponds to the automatic extraction of the data tables from HTML version of documents using tag analysis. After the data table are presented to the domain expert for validation. These table can synthesize some experimental data published in the document and so can be used to facilate the manual entering. This substep also contains the annotation of all the documents, corresponding to the to the manual semantic annotation of the selected data tables using the concepts of Biorefinery OTR. This will guide the expert in his entering task, allowing him not to forget to fulfill arguments of the selected n-ary relation concepts which guarantee the reusability of data. The n-ary relationships shape the annotation of the scientific article. In addition, these relationships are used to create annotated tables, as shown in **Figure 6a**. Here, it describes a biorefinery pretreatment process composed of a sequence of six unit operation realized in two experiments. The columns of the annotated table correspond to arguments of the relation *Milling_Solid_Quantity_Output_Relation* like *Biomass quantity*, *Treatment* (which is the type of operation) and *Total pretreatment Energy.* On the row 1 we can see that the first unit operation is a milling.

The biorefinery ontology includes three tables to be completed with annotations: Biomass composition, enzyme cocktail and process description. All the foreground data required for the next stage —establishment of the life cycle inventory — are provided in the process description table. The last substep is the storage which is automatically done by @web.



a)

Process description (figure 1) en
export / view info

n°	Output solid constituent size Unit : µm	Treatment	Experience number Unit : 1	Process step number Unit : 1	Biomass	Biomass quantity Unit : g	Total pretreatment energy Unit : kW.h.kg-1	Water quantity Unit : ml
1	[1.770e+02 ; 8.330e+02]	Milling	1.000e+00	1.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	[0.000e+00 ; inf]
2		Drying	1.000e+00	2.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	
3		Organosolv treatment	1.000e+00	3.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	1.000e+02
4		Washing and separation	1.000e+00	4.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	2.500e+01
5		Drying	1.000e+00	5.000e+00	Rice straw	5.000e+01	0.000e+00	
6		Enzymatic hydrolysis treatment	1.000e+00	6.000e+00	Rice straw	5.000e+00		
7	[1.770e+02 ; 8.330e+02]	Milling	2.000e+00	1.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	[0.000e+00 ; inf]
8		Drying	2.000e+00	2.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	
9		Organosolv treatment	2.000e+00	3.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	1.000e+02
10		Washing and separation	2.000e+00	4.000e+00	Rice straw	5.000e+01	[0.000e+00 ; inf]	2.500e+01
11		Drying	2.000e+00	5.000e+00	Rice straw	5.000e+01	0.000e+00	
12		Enzymatic hydrolysis treatment	2.000e+00	6.000e+00	Rice straw	5.000e+00		

Figure 6: *Example of @Web screenshots: (a) a process description annotation table from [54] and (b) A n-ary relationship: milling solid quantity output relationship.*

4.3 Life cycle inventory

In this stage, the life cycle inventory was done following the LCA method. The first substep the quantification of direct flows (inputs and outputs), such as electricity or acid. The foreground data from publications on rice straw and corn stover pretreatments used concerned energy, biomass, water, acid, oxidation, ionic liquids, alcohol, alkalis, buffer liquid, chemical regent and output solid. These data were grouped together in the annotation base in @Web. Once they had been extracted from @Web, the foreground data were expressed per functional unit. An LCI database is then required to associate foreground data from @Web with background data in the LCI database. Indeed, the production of electricity for biomass transformation involves the extraction and emission of matter and energy, which must be taken into account in life cycle approaches. The EcoInvent v3 database was used in this study.

The last substep is the validation of the data. Uncertainty analysis was performed by the classical method of quality evaluation based on the Weidema and Wesnaes pedigree matrix [48]. For each data item, quality was evaluated based on six parameters (reliability, exhaustively, temporal correlation, geographic correlation, technological correlation, and sample size correlation), on a scale of 1 to 6, where 1 corresponds to the highest quality and six, to the lowest. We then used a correspondence table to attribute an uncertainty factor to each evaluation [49].

In this stage, we verify the data come from @web by material balance. On the 20 articles, we choose only the 13 relevant articles where it is possible to do this balance (seven for the corn stover and six for the rice straw). Indeed, in a lot of pretreatment articles the authors do not precisely the different material use in all the process steps. These articles are listed into **Table 2** with their identification name used in the article and their type of pretreatment process described in **Section 4.2** This validation of the articles is done manually by process engineers and directly on the data from @web without to read the original article. In the end of the stage, we decided to create one experiment from each article to facilitate subsequent calculations, analysis, visualization and decision-making. This experiment was an "average experiment" from each article, by calculating mean values for all the data in each article. These mean values were then expressed per functional unit, to ensure data consistency. In our future studies, it was possible to do the LCIA on all experiments of every article but for a first study, we decide to reduce the calculation time and the visualization complexity. Indeed, some article contains more than twenty experiments.

Id.	Process type	Authors, date	Title of article
PCS1	PM-PC-EX-PS	Liu et al., 2013	Alkaline twin-screw extrusion pretreatment for fermentable sugar production [53]
PCS2	PM-PC-EX-PS	Chen et al., 2014	Screw extrude steam explosion: A promising pretreatment of corn stover to enhance enzymatic hydrolysis.
PCS3	PM-PC-PS	Liu et al., 2013	Effects of biomass particle size of steam explosion pretreatment performance for improving the enzyme digestibility of corn stover.
PCS4	PM-PC-PS	Tai et al., 2014	Impact of pretreatment with dilute sulfuric acid under moderate temperature on hydrolysis of corn stover with two enzyme systems.
PCS5	PM-PC-PS	Bals et al., 2011	Low Temperature and Long Residence Time AFEX Pretreatment of Corn Stover.
PCS6	PM-PC-PS	Zhou et al., 2014	Changes in plant cell-wall structure of corn stover due to hot compressed water pretreatment and enhanced enzymtic hydrolysis.
PCS7	PM-UFM	Lin et al., 2010	Ball Milling Pretreatment of Corn Stover for Enhancing the Efficiency of Enzymatic Hydrolysis.
PRS1	PM-PC-PS	Sheikh et al., 2013	Effect of torrefaction for the pretreatment of rice straw for ethanol production.
PRS2	PM-PC-PS	Inoue et al., 2012	Combination of hot compressed water treatment and wet disk milling for high sugar recovery yield in enzymatic hydrolysis of rice straw.
PRS3	PM-PC-UFM-PS	Inoue et al., 2012	Combination of hot compressed water treatment and wet disk milling for high sugar recovery yield in enzymatic hydrolysis of rice straw.
PRS4	PM	Ilgook et al., 2013	Effect of nitric acid on pretreatment and fermentation for enhancing ethanol production of rice straw.
PRS5	PM	Poornejad et al., 2013	Improvement of saccharification and ethanol production from rice straw by NMMO and [BMIM][OAc] pretreatments
PRS6	PM	Amiri et al., 2014	Organosolv pretreatment of rice straw for efficient acetone, butanol, and ethanol production

Table 2 : *Articles selected for the LCI in the case study*

4.4 Life cycle impact assessment

In LCA, several different methods group together different impact categories at different levels of aggregation. For this study, the ReCiPe 2008 Endpoint Hierarchical method was used. This method calculates indicators from the first level of aggregation to the final unique score. The indicators for the second level of aggregation and ReCiPe endpoint indicators for each process are presented. All the results were calculated by SimaPro software. Three LCIA graphs are proposed here : the first present the LCIA for the seven pretreatment process of corn stover (**Figure 7**), the second the LCIA for the six pretreatment process of rice straw (**Figure 8**) and the last one presents the LCIA for one type of pretreatment process - PM-PC-PS pretreatment- for the two biomass (**Figure 9**).

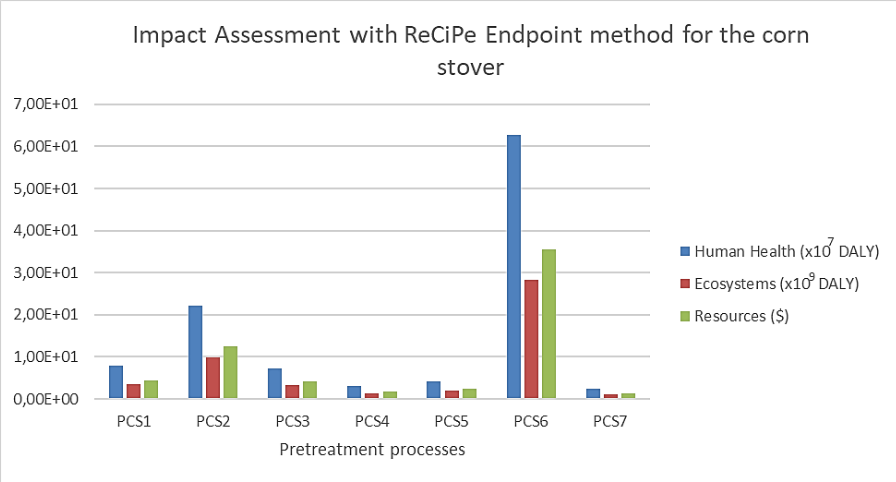


Figure 6: *Environmental Impact Assessment indicators for corn stover*

For the **Figure 6**, there are three types of pretreatment processes: PCS1 and PCS2 are PM-PC-EX-PS pretreatment, PCS3-PSC6 are PM-PC-PS pretreatment and PCS7 is PM-UFM pretreatment. Firstly, we can conclude the mechanical pretreatment is the least misleading environmental impacts. Secondly, between the PM-PC-PS pretreatment processes there are many differences – mostly between PSC6 and the others. Thirdly, the PM-PC-EX-PS so the PCS1 and PCS2 have more impacts than other pretreatments if we make an exception of the process PCS6. To explicate the PCS6 exception we wait to see the comparison of rice straw and corn stover on the same type of processes.

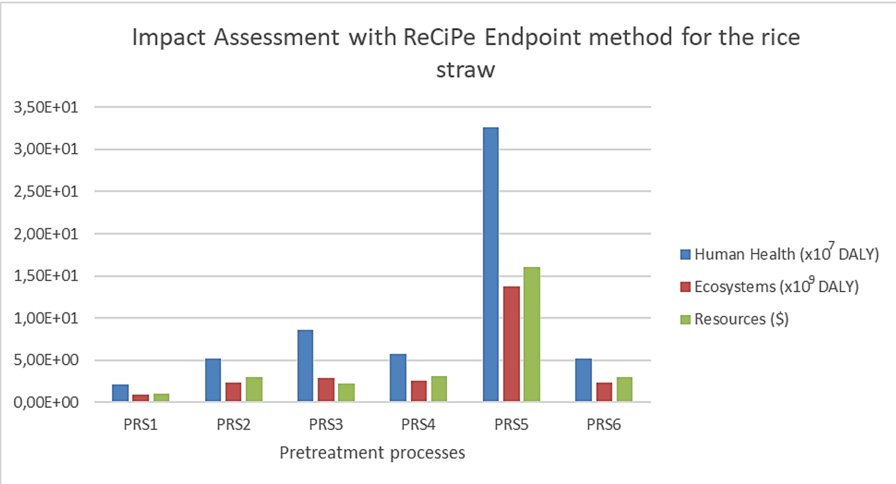


Figure 7: *Environmental Impact Assessment indicators for corn stover*

For **Figure 7**, there are also three types of pretreatment processes: PRS1 and PRS2 are PM-PC-PS pretreatment, PRS3 is PM-PC-UFM-PS pretreatment and PRS4-PRS6 are PM pretreatment. Firstly, we can observe the PRS5 which has higher impacts. We go back up the calculation and we discover a high water quantity which improves the impacts. The impacts of the other process are substantially the same and we cannot highlight clusters.

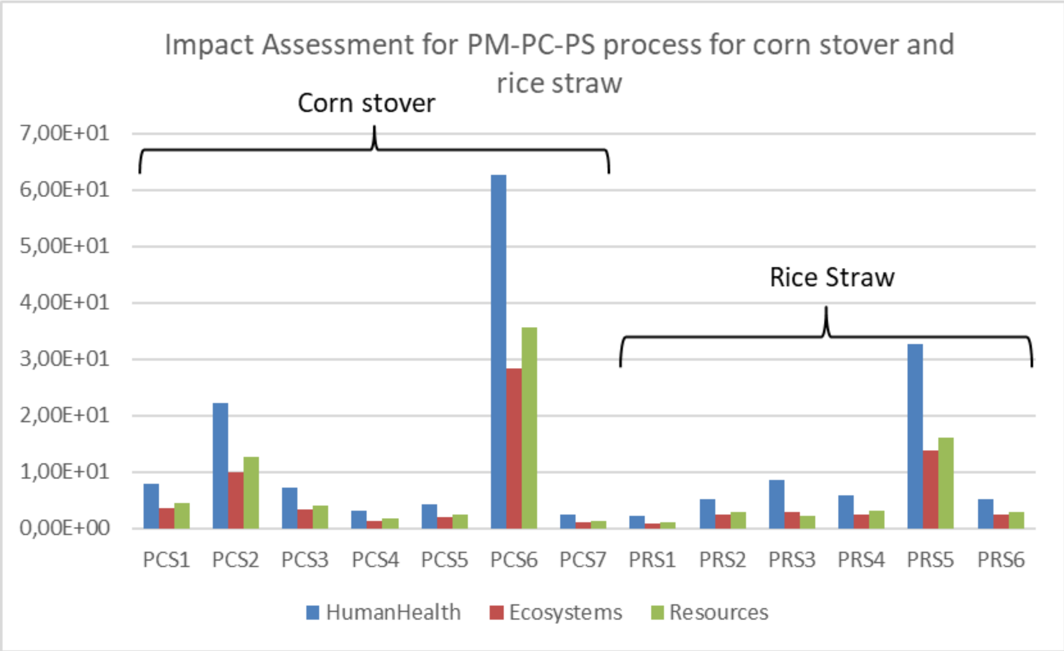


Figure 8: *Environmental impact assessment indicators for PM-PC-PS processes with corn stover and rice straw.*

For **Figure 8**, the process PCS6 has very high impacts unlike other processes of the same type. This result can be surprising and it is necessary to verify the results by going back up the calculations to the foreground data. This verification highlights a large use of water (almost 10 times higher than other processes) and alkaline (here ammonium sulfate). Recall that all these results are done by functional unit so can be compared.

5 Conclusion and perspective

In this study, we developed a new approach, in which intensive data and knowledge methods are used to complete data for sustainability analysis. This approach involves five stages: determination of the goal and scope of the study, the structuration and extraction of intensive data from heterogeneous data sources, establishment of the life cycle inventory, impact sustainability assessment and, finally, the analysis and interpretation of the results. Like the LCA method, this methodology is generic, because the ontologies established for exploitation of the foreground data are generic. This coupling of intensive data and knowledge method and sustainability assessment can be applied to all processes. The intensive data and knowledge method utilized in this paper is the KE method, it is possible to use other methods following the data being processed. In this study, we use the example of a biorefinery. The proposed pipeline should enable researchers, and other users, to identify the "best" process for a specific biomass. Environmental and economic indicators can be used. The addition of document data like reliability score can be used to complete the results or set aside data from an article with low reliability score. In the case study, only environmental indicators are shown. The main result is that the mechanical pretreatment processes of the biomass are the processes with the lowest ecological impacts. However, these results have to be completed because of the lack of energy data in the article that is used to assess the foreground system. This approach could be improved by enhancing ranking procedures and including scaling (to address industry needs). Furthermore, the environmental indicators could be combined with economic indicators, providing a more general overview of different processes and biomasses. The approach stays a first approach of the intensive data and knowledge-driven approach for sustainability analysis that will be completed with economic and social assessments and more advanced big data methods.

Acknowledgments

This project was funded by the 3BCAR French Carnot network (http://www.3bcar.fr), which brings together seven French research units (500 researchers) involved in the design of processes for transforming biomass into bioenergy, biomaterials and biomolecules.

References

1. Ellen Macarthur Foundation: Towards a circular economy: business rationale for an accelerated transition. Ellen Macarthur Foundation (2015)
2. Ghisellini, P., Cialani, C., Ulgiati, S.: A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. J. Clean. Prod. 114, 11–32 (2016). https://doi.org/10.1016/j.jclepro.2015.09.007
3. Belaud, J.-P., Adoue, C., Vialle, C., Chorro, A., Sablayrolles, C.: A circular economy and industrial ecology toolbox for developing an eco-industrial park: perspectives from French policy. Clean Technol. Environ. Policy. 21, 967–985 (2019). https://doi.org/10.1007/s10098-019-01677-1
4. Millati, R., Cahyono, R.B., Ariyanto, T., Azzahrani, I.N., Putri, R.U., Taherzadeh, M.J.: Agricultural, Industrial, Municipal, and Forest Wastes. In: Sustainable Resource Recovery and Zero Waste Approaches. pp. 1–22. Elsevier (2019)
5. Garnett, T., Appleby, M.C., Balmford, A., Bateman, I.J., Benton, T.G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., Herrero, M., Hoffmann, I., Smith, P., Thornton, P.K., Toulmin, C., Vermeulen, S.J., Godfray, H.C.J.: Sustainable Intensification in Agriculture: Premises and Policies. Science. 341, 33–34 (2013). https://doi.org/10.1126/science.1234485
6. West, P.C., Gerber, J.S., Engstrom, P.M., Mueller, N.D., Brauman, K.A., Carlson, K.M., Cassidy, E.S., Johnston, M., MacDonald, G.K., Ray, D.K., Siebert, S.: Leverage points for improving global food security and the environment. Science. 345, 325–328 (2014). https://doi.org/10.1126/science.1246067
7. Horton, P., Koh, L., Guang, V.S.: An integrated theoretical framework to enhance resource efficiency, sustainability and human health in agri-food systems. J. Clean. Prod. 120, 164–169 (2016). https://doi.org/10.1016/j.jclepro.2015.08.092
8. Food SCP Round Table European Commission: Continous Environmental Improvement - Final Report. (2012)
9. Raymond, R.: Improving food systems for sustainable diets in a green economy. 2012 United Nations Conference on Sustainable Development: Governance for Greening the Economy with Agriculture (2012)
10. Davis, C.B., Aid, G., Zhu, B.: Secondary Resources in the Bio-Based Economy: A Computer Assisted Survey of Value Pathways in Academic Literature. Waste Biomass Valorization. 8, 2229–2246 (2017). https://doi.org/10.1007/s12649-017-9975-0
11. Joglekar, S.N., Tandulje, A.P., Mandavgane, S.A., Kulkarni, B.D.: Environmental Impact Study of Bagasse Valorization Routes. Waste Biomass Valorization. 10, 2067–2078 (2019). https://doi.org/10.1007/s12649-018-0198-9
12. Chuetor, S., Luque, R., Barron, C., Solhy, A., Rouau, X., Barakat, A.: Innovative combined dry fractionation technologies for rice straw valorization to biofuels. Green Chem. 17, 926–936 (2015). https://doi.org/10.1039/C4GC01718H
13. Barakat, A., de Vries, H., Rouau, X.: Dry fractionation process as an important step in current and future lignocellulose biorefineries: A review. Bioresour. Technol. 134, 362–373 (2013). https://doi.org/10.1016/j.biortech.2013.01.169
14. Zhu, J.Y., Pan, X., Zalesny, R.S.: Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance. Appl. Microbiol. Biotechnol. 87, 847–857 (2010). https://doi.org/10.1007/s00253-010-2654-8
15. ISO 14044:2006: Environmental management - Life cycle Assessment - Requirements and guidelines. International Organization for Standardization, Geneva, Switzerland (2006)
16. ISO 14040:2006: Environmental management - Life cycle Assessment - Principles and Framework. International Organization for Standardization, Geneva, Switzerland (2006)
17. Burgess, A.A., Brennan, D.J.: Application of life cycle assessment to chemical processes. Chem. Eng. Sci. 56, 2589–2604 (2001). https://doi.org/10.1016/S0009-2509(00)00511-X
18. Elghali, L., Clift, R., Sinclair, P., Panoutsou, C., Bauen, A.: Developing a sustainability framework for the assessment of bioenergy systems. Energy Policy. 35, 6075–6083 (2007). https://doi.org/10.1016/j.enpol.2007.08.036

19. Clift, R., Frischnecht, R., Huppes, G., Tillman, A.-M., Wlidema, B.: Toward a Coherent Approach to Life Cycle Inventory Analysis. Report of the Working Group on Inventory Enhancement, Brussels (1998)

20. Guinée, J.B. ed: Handbook on life cycle assessment: operational guide to the ISO standards. Kluwer Academic Publishers, Dordrecht ; Boston (2002)

21. Suh, S., Huppes, G.: Methods for Life Cycle Inventory of a product. J. Clean. Prod. 13, 687–697 (2005). https://doi.org/10.1016/j.jclepro.2003.04.001

22. Snijders, C., Matzat, U., Reips, U.-D.: “Big Data”: Big Gaps of Knowledge in the Field of Internet Science. International Journal of Internet Science. 7 (1), 1–5 (2012)

23. Doan, A., Halevy, A., Ives, Z.G.: Principles of data integration. Morgan Kaufmann, Waltham, MA (2012)

24. Noy, N.F.: Semantic integration: a survey of ontology-based approaches. ACM SIGMOD Rec. 33, 65 (2004). https://doi.org/10.1145/1041410.1041421

25. Hao, Y., Zhu, X., Huang, M., Li, M.: Discovering patterns to extract protein-protein interactions from the literature: Part II. Bioinformatics. 21, 3294–3300 (2005). https://doi.org/10.1093/bioinformatics/bti493

26. Buche, P., Dibie-Barthelemy, J., Ibanescu, L., Soler, L.: Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. IEEE Trans. Knowl. Data Eng. 25, 805–819 (2013). https://doi.org/10.1109/TKDE.2011.245

27. Le Minh, Q., Truong, S.N., Bao, Q.H.: A Pattern Approach for Biomedical Event Annotation. Presented at the Proceedings of BioNLP Shared Task 2011 Workshop , Portland, Oregon, USA (2011)

28. Zhou, D., Zhong, D., He, Y.: Biomedical Relation Extraction: From Binary to Complex. Comput. Math. Methods Med. 2014, 1–18 (2014). https://doi.org/10.1155/2014/298473

29. Dibie, J., Dervaux, S., Doriot, E., Ibanescu, L., Pénicaud, C.: [MS]^{2}O – A Multi-scale and Multi-step Ontology for Transformation Processes: Application to Micro-Organisms. In: Haemmerlé, O., Stapleton, G., and Faron Zucker, C. (eds.) Graph-Based Representation and Reasoning. pp. 163–176. Springer International Publishing, Cham (2016)

30. Rijgersberg, H., Wigham, M., Top, J.L.: How semantics can improve engineering processes: A case of units of measure and quantities. Adv. Eng. Inform. 25, 276–287 (2011). https://doi.org/10.1016/j.aei.2010.07.008

31. Xu, M., Cai, H., Liang, S.: Big Data and Industrial Ecology: Big Data and Industrial Ecology. J. Ind. Ecol. 19, 205–210 (2015). https://doi.org/10.1111/jiec.12241

32. Cooper, J., Noon, M., Jones, C., Kahn, E., Arbuckle, P.: Big Data in Life Cycle Assessment: Big Data in Life Cycle Assessment. J. Ind. Ecol. 17, 796–799 (2013). https://doi.org/10.1111/jiec.12069

33. Bhinge, R., Srinivasan, A., Robinson, S., Dornfeld, D.: Data-intensive Life Cycle Assessment (DILCA) for Deteriorating Products. Procedia CIRP. 29, 396–401 (2015). https://doi.org/10.1016/j.procir.2015.02.192

34. Zhang, Y., Luo, X., Buis, J.J., Sutherland, J.W.: LCA-oriented semantic representation for the product life cycle. J. Clean. Prod. 86, 146–162 (2015). https://doi.org/10.1016/j.jclepro.2014.08.053

35. Zhou, L., Zhang, C., Karimi, I.A., Kraft, M.: An ontology framework towards decentralized information management for eco-industrial parks. Comput. Chem. Eng. 118, 49–63 (2018). https://doi.org/10.1016/j.compchemeng.2018.07.010

36. Simperl, E., Bürger, T., Hangl, S., Wörgl, S., Popov, I.: ONTOCOM: A reliable cost estimation method for ontology development projects. J. Web Semant. 16, 1–16 (2012). https://doi.org/10.1016/j.websem.2012.07.001

37. Lee, C.-S., Wang, M.-H., Yan, Z.-R., Lo, C.-F., Chuang, H.-H., Lin, Y.-C.: Intelligent estimation agent based on CMMI ontology for project planning. Presented at the October (2008)

38. Zhang, J., Li, H., Zhao, Y., Ren, G.: An ontology-based approach supporting holistic structural design with the consideration of safety, environmental impact and cost. Adv. Eng. Softw. 115, 26–39 (2018). https://doi.org/10.1016/j.advengsoft.2017.08.010

39. Shibasaki, M., Fischer, M., Barthel, L.: Effects on Life Cycle Assessment — Scale Up of Processes. In: Takata, S. and Umeda, Y. (eds.) Advances in Life Cycle Engineering for Sustainable Manufacturing Businesses. pp. 377–381. Springer London, London (2007)

40. Kopnina, H.: Green-washing or best case practices? Using circular economy and Cradle to Cradle case studies in business education. J. Clean. Prod. 219, 613–621 (2019). https://doi.org/10.1016/j.jclepro.2019.02.005

41. Lousteau-Cazalet, C., Barakat, A., Belaud, J.-P., Buche, P., Busset, G., Charnomordic, B., Dervaux, S., Destercke, S., Dibie, J., Sablayrolles, C., Vialle, C.: A decision support system for eco-efficient biorefinery process comparison using a semantic approach. Comput. Electron. Agric. 127, 351–367 (2016). https://doi.org/10.1016/j.compag.2016.06.020

42. Destercke, S., Buche, P., Charnomordic, B.: Evaluating Data Reliability: An Evidential Answer with Application to a Web-Enabled Data Warehouse. IEEE Trans. Knowl. Data Eng. 25, 92–105 (2013). https://doi.org/10.1109/TKDE.2011.179

43. Touhami, R., Buche, P., Dibie-Barthélemy, J., Ibănescu, L.: An Ontological and Terminological Resource for N-ary Relation Annotation in Web Data Tables. In: Proceedings of the 2011th Confederated International Conference on On the Move to Meaningful Internet Systems - Volume Part II. pp. 662–679. Springer-Verlag, Berlin, Heidelberg (2011)

44. ISO/TS 14048:2002: Environmental management — Life cycle assessment — Data documentation format. International Organization for Standardization, Geneva, Switzerland (2002)

45. Frischknecht, R., Jungbluth, N., Althaus, H.-J., Doka, G., Dones, R., Heck, T., Hellweg, S., Hischier, R., Nemecek, T., Rebitzer, G., Spielmann, M.: The ecoinvent Database: Overview and Methodological Framework (7 pp). Int. J. Life Cycle Assess. 10, 3–9 (2005). https://doi.org/10.1065/lca2004.10.181.1

46. U.S. Life Cycle Inventory Database, https://www.nrel.gov/lci/

47. European Commission: International Reference Life Cycle Data System (ILCD) Handbook - General guide for Life Cycle Assessment - Provisions and Action Steps, https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/international-reference-life-cycle-data-system-ilcd-handbook-general-guide-life-cycle

48. Huijbregts, M.A.J., Steinmann, Z.J.N., Elshout, P.M.F., Stam, G., Verones, F., Vieira, M., Zijp, M., Hollander, A., van Zelm, R.: ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. Int. J. Life Cycle Assess. 22, 138–147 (2017). https://doi.org/10.1007/s11367-016-1246-y

49. Sala, S., Wolf, M.-A., Pant, R.: Characterisation factors of the ILCD Recommended Life Cycle Impact Assessment methods. Database and supporting information. (2012)

50. SimaPro | The world's leading LCA software, https://simapro.com/

51. Spatari, S., Betz, M., Florin, H., Baitz, M., Faltenbacher, M.: Using GaBi 3 to perform life cycle assessment and life cycle engineering. Int. J. Life Cycle Assess. 6, 81 (2001). https://doi.org/10.1007/BF02977842

52. Bouyssou, D. ed: Evaluation and decision models with multiple criteria: stepping stones for the analyst. Springer, New York (2006)

53. Liu, C., van der Heide, E., Wang, H., Li, B., Yu, G., Mu, X.: Alkaline twin-screw extrusion pretreatment for fermentable sugar production. Biotechnol. Biofuels. 6, 97 (2013). https://doi.org/10.1186/1754-6834-6-97

54. Amiri, H., Karimi, K., Zilouei, H.: Organosolv pretreatment of rice straw for efficient acetone, butanol, and ethanol production. Bioresour. Technol. 152, 450–456 (2014). https://doi.org/10.1016/j.biortech.2013.11.038