Big data for agri-food 4.0: Application to sustainability management for by-products supply chain

Jean-Pierre Belaud[a,*], Nancy Prioux[a], Claire Vialle[b], Caroline Sablayrolles[b]

[a] *Laboratoire de Génie Chimique, Université de Toulouse, CNRS, Toulouse, France*
[b] *Laboratoire de Chimie Agro-industrielle, Université de Toulouse, INRA, Toulouse, France*

A R T I C L E I N F O

Keywords:
Industry 4.0
Sustainability
Circular economy
Biorefinery
Agriculture supply chain
Life-cycle assessment

A B S T R A C T

The bioconversion of lignocellulosic biomass is a promising method for the production of bio-energy, biomolecules and biomaterials. Pretreatment of the lignocellulosic biomass is an essential step in this process. The choice of pretreatment process is a difficult one, and there are currently no clear criteria on which to base this choice. This project, with its sustainability and agri-food perspective, used environmental impacts to assess the various processes and their panels of technologies. The approach developed integrates big data, to improve sustainability management in supply chain design, with the aim of valorising agricultural waste. In five main steps, this approach combines concepts from industry 4.0, sustainability and the agri-food industry. We apply this approach to a case study in the domain of agricultural waste valorisation: the pretreatment of lignocellulosic biomass in the rice supply chain.

1. Introduction and background

Since the beginning of the 1970s, human influence on the Earth and its resources — through the economic, scientific and technological development of our industrial society — has steadily increased, resulting in an ever-greater impact on the environment. Awareness of these ecological problems has sparked new ideas for more eco-friendly development. The Brundtland report (1987) marked the start of "sustainable development", with the first definition of this term as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs." Sustainable development thus has three main objectives: economic efficiency, social fairness and environmental sustainability. It requires profound changes in the way we think, design and use resources, and in economic and social structures for consumption and production patterns. Life-cycle thinking can help to improve environmental performances, and social and economic benefits can be derived from approaches taking the full life-cycle of the agricultural supply chain into account. Indeed, this approach can be used to minimise the impact in some areas, whilst preventing further impacts in other areas. Sustainable development requires the stable simultaneous consideration of economic, environmental and social aspects (Hardaker, 1997). A new business model for more sustainable development has recently emerged: the circular economy (CE) (Mathews and Tan, 2011). This model helps to reconcile economic, environmental and social aspects. Ghisellini et al. (2016) have reviewed scientific articles on CE and have discussed the origins of CE and the principles and limitations of CE models. In France, the circular economy was included in the nine areas of the SNTEDD (National Ecological Transition Strategy for Sustainable Development) proposed by the French government in 2015 (Belaud et al., 2019). The term "ecological transition" is used to describe the shift towards more sustainable development. This transition is a major societal concern and will require broad knowledge and skills (Reichman et al., 2011) from many different sciences (Palmer et al., 2005).

The digital transformation of current societies has led to a new era in industry, which can be described as "industry 4.0", which may facilitate the ecological transition. Industry has changed considerably since its beginnings in the 18th century. Four industrial revolutions are now considered to have occurred.

* Corresponding author.
E-mail addresses: jeanpierre.belaud@ensiacet.fr (J.-P. Belaud), nancy.prioux@ensiacet.fr (N. Prioux), claire.vialle@ensiacet.fr (C. Vialle), caroline.sablayrolles@ensiacet.fr (C. Sablayrolles).

https://doi.org/10.1016/j.compind.2019.06.006

Industrial revolution 1.0 corresponds to the introduction of machinery powered by the local generation of steam, which uncoupled production from the limitations of human manual effort. Industry 2.0 corresponds to the period after 1870 in which “scientific management” or Taylorism introduced the division of labour, assembly lines and the use of electric energy machines. The 20th century, with the development of electronics, computing and robotic manufacturing, ushered in the era of industry 3.0, which focused on quality and cost performances. Automation has provided opportunities to optimise manufacturing processes and improve productivity through the design of more flexible, ergonomic and safer machinery. Industry 4.0, based on the technological concepts of cyber physical systems (CPS) (Babiceanu and Sekr, 2016), internet of thing (IoT), big data and an internet of services (Kagermann et al., 2013), is an umbrella term for the technologies and concepts of value chain organisation, which facilitates the development of “smart factories”. One of the core principles of industry 4.0 is data management, from collection to analysis, and the integration of information technologies, manufacturing and operation systems as a way to acquire data in a more timely, rapid and flexible manner (Brettel et al., 2014). This transformation from industry 3.0 to industry 4.0 is often referred to as the “digitalisation process” or “digital transition”. Big data and related methods and tools form one of the pillars supporting this transformation (Chen and Zhang, 2014).

Big data are information assets with a high volume, velocity and variety (De Mauro et al., 2016), making them difficult to manage with common tools (Hampton et al., 2013). Big data technologies can be applied to the CE and industrial ecology. Ming et al. (2015) explored the possible contribution of big data to industrial ecology through several examples combining these two domains. Knowledge engineering (KE) is a technique from knowledge-based systems that can now be applied to big data. Life-cycle thinking would clearly benefit from the combination of the huge amounts of data now available with KE techniques for their exploitation. This would make it possible to obtain additional and surrogate data in situations in which specific data cannot be collected, rather than having to rely on default and missing values. This approach requires a set of hypotheses. The main goal of KE is to structure knowledge into formal representations for exploitation by computers. This structuring of data is particularly important when handling large amounts of data, which push standard statistical software to its limits (Snijders et al., 2012). KE methods used a standardised vocabulary to structure the experimental data and their meaning. This structuring may be based on an ontology representing the experimental data of interest (Noy, 2004). Ontologies are knowledge representation models that can be used to link data and to provide automated tools for reasoning (Doan et al., 2012). Once the data have been structured into ontologies, they can be homogenised and used to define and calculate criteria for the assessment of processes (Liao et al., 2015). However, only a few studies have explored the application of this approach in this domain. Belaud et al. (2019) designs intensive data and information systems to manage sustainable development in the frame of eco-industrial area. Cooper et al. (2013) used big data to complete the background data. The background system consists of all other processes interacting directly with the foreground system. “Data-intensive life-cycle assessment” (Bhinge et al., 2015) uses KE-based approaches to adapt life-cycle assessment (LCA) methods, incorporating technological developments that may modify LCA results for a given product over time. Finally, big data and KE can be used to represent the life cycle of a product or service (Zhang et al., 2015), with all the intermediate flows, emissions and extractions, in ontology-based LCA.

Europe generates more than 700 million tonnes of agricultural waste annually (Pavwelczyk, 2005). The projected increase in the world’s population will undoubtedly be accompanied by increases in waste production and its impact on the environment. In addition, human activities are decreasing the availability of agricultural land, with inevitable impacts on agricultural systems. New agricultural technologies should facilitate sustainable intensification, the “best” approach for the future of agriculture (Garnett et al., 2013). However, this intensification will lead to more waste of products and resources (West et al., 2014). According to Horton et al. (2016), the parametrisation of waste in agriculture is a major challenge in attempts to achieve sustainability. We can identify two classes of waste: waste from inputs, such as water or fertiliser, and waste due to the incomplete conversion or processing of materials in the supply chain, from crop production to food consumption. This second class of agricultural waste includes lignocellulosic by-products, corresponding to the essential structural compounds of the cell walls of lignified plants. Lignocellulosic biomass is one of the most abundant and cheapest renewable resources on Earth. Its bioconversion is a promising approach to the production of bioenergy, biomolecules and biomaterials. This process involves enzymatic hydrolysis of the biomass to release glucose. The lignocellulosic biomass has four main components: cellulose, hemicellulose, lignin, and phenolic acids. Only two of these components, cellulose and hemicellulose, can be hydrolysed to generate glucose. For overall sustainability, the processes used to generate these bioresources must be sustainable. Assessments of sustainability are increasing being incorporated into processes in the agro-food domain (Food SCP Round Table European Commission, 2012; Raymond, 2012). Wolferd et al. explored smart farms through farm management, farm processes, network management organisation and network management technology(Wolfert et al., 2017). A few studies since 2010 have focused on climate change and big data in the domain of agriculture (Pivoto et al., 2018). An integrated theoretical framework developed by Horton et al. (2016) considered the agri-food supply chain — from land to people — by integrating big data and data for all actors with any influence on agri-food businesses. A generic agri-food ecosystem template was created, including the key actors, external influences, components and impacts.

We focused on the valorisation of agricultural waste and by-products, considering the ways in which big data could support the management of sustainability. The pretreatment of lignocellulosic biomass before its enzymatic hydrolysis is essential, to ensure good yields. Various pretreatment methods have been studied in detail over the last 30 years. However, it remains difficult to choose between these different processes in terms of the available biomass and product quality, and criteria are lacking to guide this choice. We studied the relationships between product quality and biomass pretreatment. Focusing on sustainability, we compared the various pretreatment processes and the technologies involved in terms of their economic and environmental impacts. Environmental impact can be assessed in various ways. We used the life-cycle assessment (LCA) method, because this approach is truly systemic and based on life-cycle thinking. LCA evaluates environmental aspects and potential environmental impacts throughout the life cycle of a product or process (ISO, 2006). We initially extracted relevant information from heterogeneous sources for the analysis of pretreatments for lignocellulosic biomass valorisation. Our approach combines concepts from industry 4.0, sustainability and the agri-food industry. In particular, it incorporates big data, to improve sustainability management in supply chain design for the valorisation of agriculture wastes and outputs of agri-food applications. In the next section, we provide an overview of our research approach and the associated detailed workflow. In section 3, we apply this approach to a case study: sustainability management for four agri-food processes, guiding decisions relating to the supply chain and technologies in rice production. We then discuss the conclusions of this study and future perspectives.

2. Big data for the agriculture by-product supply chain

2.1. Materials and methods

Big data can be used at various levels of sustainability management. One of the challenges in by-product valorisation in the agricultural supply chain is designing the "best" process. The by-product valorisation supply chain includes several operational stages, from *biomass choice* to *waste disposal*, and it passes through various *transformation stages* and *upstream/downstream processes*. Each stage can be defined with impact methods and indicators relating to three areas: *economic, environmental* and *social*. Once the various stages have been described, the researcher can choose the biomass and the most appropriate process with the decision support tool, which takes into account all the indicators in each area. The main goal of this approach is to analyse the different valorisation systems and provide support for *group-based decision-making*. The link between the decision support system and the various data makes it possible to treat various types of data whilst maintaining a high-throughput for big data processing. The various data used are listed in Fig. 1. *Public web* data are available to anyone with a web browser and include weather data, world prices for raw materials and impact factors for LCA methods (ReCiPe for example). *Corporate data* are data obtained from companies at any stage of their activity (from the setting up of the company until its closure). In the case of agricultural by-product valorisation, the "company" is a biorefinery, and each biorefinery has its own transformation process data. *Field data* are data describing biomass quality (cellulose content) and mass, and moisture content, for example. *Technological data* relate to the technologies used in valorisation processes, such as cutting and milling technologies, inputs, rotation speeds and energy values, for example. *LCA databases* are databases widely used in LCA, such as EcoInvent and Gabi. *Scientific databases* are databases of scientific articles. In our example (Section 3), the articles were obtained from Web of Science and Science Direct.

We have three main axes of interest (Fig. 1). The workflow to support the link between big data and sustainability assessments for valorisation of the agricultural by-product supply chain is detailed in Section 2.2. This approach was adapted to the *agri-food industry* by making use of concepts from *industry 4.0* and *sustainability* management. In particular, we retained the *Big Data* pillar from *Industry 4.0* and *sustainability assessment* from *sustainability* management. Fig. 1 illustrates one path of digital transformation based on the integration of big data into the agri-food industry. The valorisation of the agricultural by-products supply chain can be split into five elements: *lignocellulosic biomass*, *transformation processes and technologies*, *inputs and outputs*, *products and wastes,* and *upstream and downstream processes*. Each of these categories can be described with heterogeneous data and can influence another category. For example, the type of biomass can influence the type of transformation processes. All data are of importance and influence the social, economic and environmental indicators. In the sustainability of agricultural by-product valorisation, the goal is to integrate all the data into process design, but this is very difficult. For example, for environmental assessments complying with life-cycle thinking, process data are required. Obtaining such data is time-consuming and requires expensive experiments. Alternatively, data can be obtained from scientific publications and other sources, with the automatic or manual use of big data. It is, indeed, possible to make use of these data and, therefore, to obtain foreground data for sustainability analysis, whereas background data are generally available from the *LCA database*.

The integration of *sustainability assessment* into the *Agri-Food Industry* will facilitate the ecological transition. Our approach is divided into five major steps (Fig. 2): *goal*, *data architecture*,

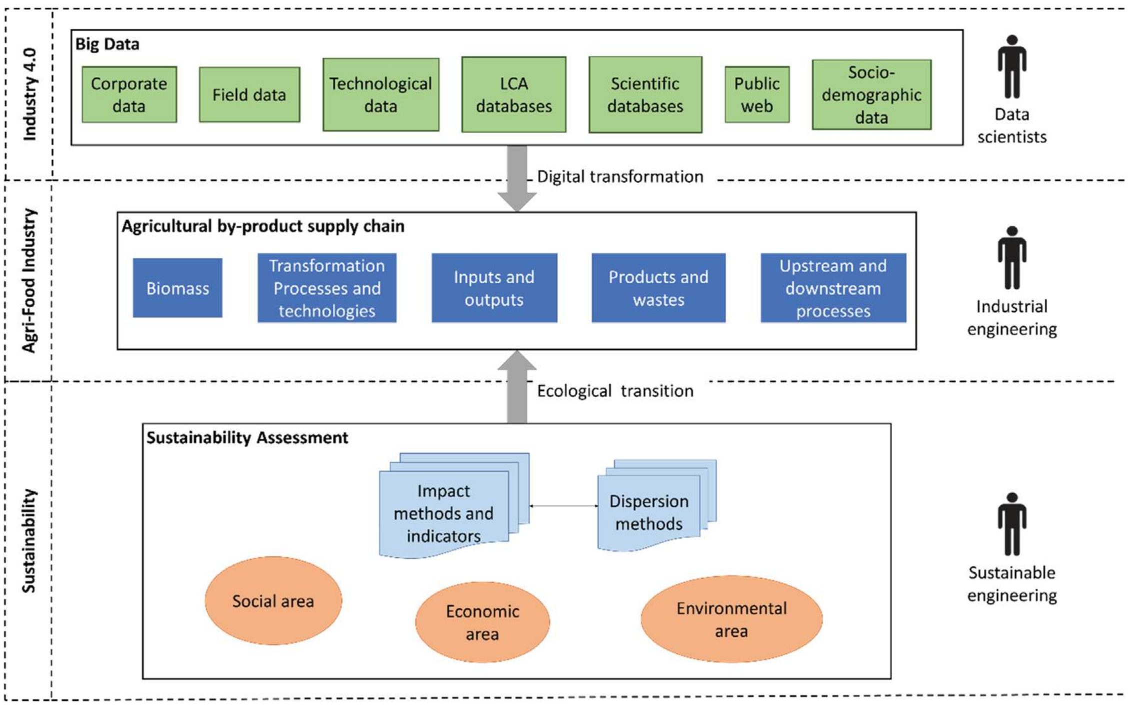


Fig. 1. Big data and sustainability assessment for agricultural by-product valorization.

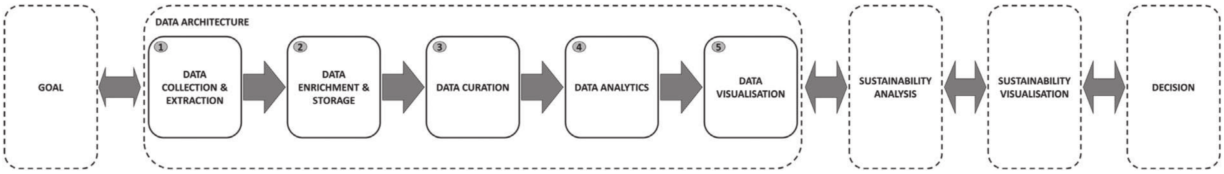


Fig. 2. The various steps in the approach.

sustainability analysis, *sustainability visualisation* and *decision*. Each step has its own substeps and passage to and fro between the various steps is recommended. The second step of the proposed methodology concerns the construction of the big data architecture: *data collection and extraction, data enrichment and storage, data curation, data analytics* and *data visualisation* (Levin et al., 2013).

There are five substeps in the construction of the data architecture: *data collection and extraction*, *data enrichment and storage, data curation, data analytics* and *data visualization*. Chen et al. (2012) compared these substeps to Business Intelligence and Analytics. The data come from various heterogeneous sources and can be structured or unstructured. The tools commonly used for ETL (extraction, transformation and loading) in data management and warehousing are unsatisfactory in this context. Specific methods and tools are therefore required.

The *collection and extraction* of data from structured databases require specific tools, such as data queries (SQL requests) or online analytical processing (OLAP). This substep is more complicated for unstructured data from the web. We extract the various webpages containing the data of interest. The metadata associated with these webpages can be used for their classification and to provide access to their content. For example, the definition of the MARC format in the early 1960's normalised documentary resource metadata. Thanks to Linked Web development including RDF (Resource Description Framework, a standard from W3C) in particular, it is possible to use SPARQL to request the RDF. The extraction process generates a structured table and differs according to the format of the webpages: API, HTML or pdf. Extraction may be automatic, semi-automatic or manual. Data can be extracted automatically from webpages with an API format. However, the extraction of the relevant data from scientific articles requires a reading guide, which is generated from ontologies and pretreatment expert analysis. This type of extraction is, thus, semi-automatic. In the *data enrichment and storage* substep, the extracted data are stored in relational database management systems (DBMS). Enrichment is the process of adding data to the DBMS. These data come from experts or models. The models are empiric numerical simulations of unit operation type, thermodynamics and energy for the control of flows, transformations and transfers.

The *data curation* substep involves the cleaning up, addition and deletion of data for the management of volume and value. Following this curation, a second, more accurate and accessible database can be generated. Curation saves time in subsequent substeps and prevents incorrect interpretation during the *data analytics* substep and the *sustainability analysis* step. The *data analytics* substep depends on the goal, the data domains and the decision-makers. For types of analysis are possible: descriptive analysis (what happened?), diagnostic analysis (why did it happen?), predictive analysis (what will happen?) and prescriptive analysis (how can we make it happen?). Descriptive and diagnostic analyses make use of a number of different methods and algorithms, such as summarisation, standard deviation, linear or non-linear dependences, factor analysis and classification methods (decision tree induction, Bayesian networks, k-nearest neighbour classifier). Some of these methods are visual, and *data visualisation* may therefore be included in the *data analytics* substep. *Data visualization* may also be achieved by plotting the raw data as a simple or interactive graph.

2.2. Detailed workflow

Fig. 3 shows the different steps supporting the approach described in Fig. 2. In the first step, *goal*, system boundaries must be clearly defined, and life-cycle thinking (LCT) is recommended-for this purpose. LCT encourages a "from cradle to grave" or "from cradle to cradle" approach. In the CE model, the part of the life cycle in which the product is used is a key element for progress towards ecological transition. "From cradle to gate" approaches are oftenpreferred because the integration of downstream elements into sustainability analyses can be tedious and difficult. In particular, scientists and engineers often find it impossible to take the behaviour of end-users or consumers into account in their models. System boundaries have a strong effect on the assessments subsequently performed. For example, it must be specified whether the upstream biomass supply chain is taken into account. Once the goal and scope have been defined, the supply chain, technologies and transformation processes must be described. This description must be as complete as possible in terms of the process operations, the study location, the various inputs and outputs, and the type of energy used, for example. These details ensure the pertinence of the data collected. The last substep is choosing whether to study economic, social or environmental items. Studies may deal with one, two or all three of these areas. For complete sustainability management, all three areas should be included, but social elements are often removed from the analysis due to methodological limitations and time and cost concerns.

In the second step, *Data Architecture*, the five numbered substeps on Fig. 2 are similar to common steps in big dataanalysis. The data to be used must first be chosen. This requires several questions to be addressed: Where do I find the data? What kind of data? What are the uncertainties on the data? What data do I already have? What degree of data automation do I need? The list of different data in Fig. 1 is not exhaustive, and other types of data can be added, depending on the goal of the project. In this step, KE methods can be useful for collecting data. The data can be extracted in various ways (e.g. CSV, SQL, HTML, XML). Moreover, the integration of big data into agricultural supply chain valorisation does not always involve automatic processing. For semi-structured bases, such as scientific databases, experts from the domain concerned must select and verify the data exported. For structured bases, such as LCA databases, the data are automatically exported to the decision support system. The *data enrichment and storage* substeps differ between studies of different complexities or with the degree of

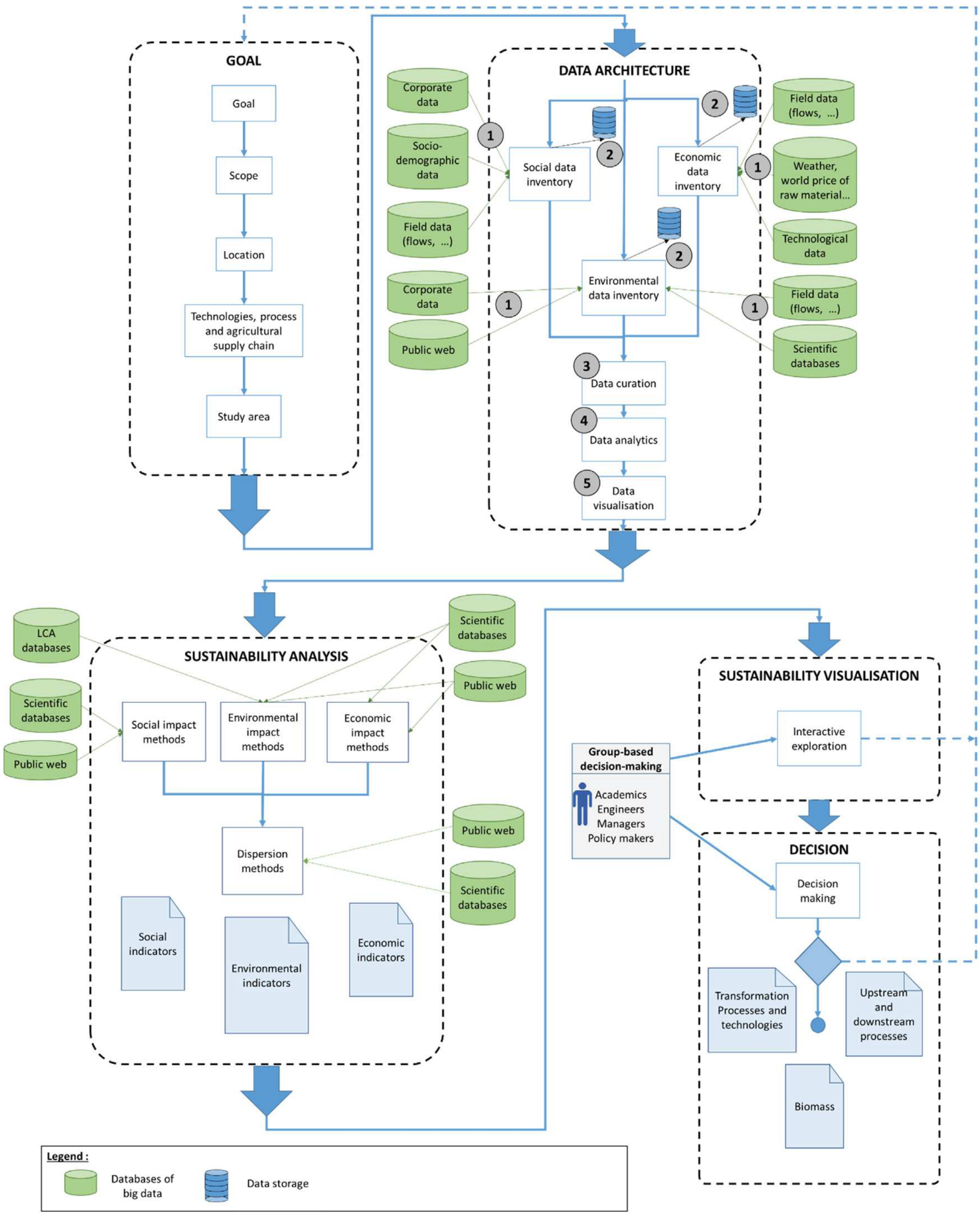


Fig. 3. Workflow of approach.

automisation. *Data curation* involves the cleaning up of the data, the addition of expert data and the deletion of abnormal data. Finally, *data analytics* and the *data visualisation* depend on the goal, the data and the decision-makers.

The third step, *sustainability analysis*, involves choosing the impact methods, the indicators and the dispersion methods in accordance with each area of sustainability management. This choice is based on the two previous steps. The most important question is whether the methods used are appropriate for the limits and the data chosen. The choice of method depends on the location at which the study takes place, the type of data and the limits of the study. It is important to check that all the hypotheses of each method are satisfied before applying the method. Each method belongs to a particular category and has its own limitations. The method may be based on a single criterion (carbon assessment) or multi-criteria (LCA), qualitative or quantitative, "product"-oriented or "organisation"-oriented.

Sustainability visualisation is the fourth step. Different types of visualisation may be used, depending on the group-baseddecision-making process, the goal of the study or the choice of analysis method, but this visualisation should never be ignored (Belaud et al., 2014). The last step is *Decision*. This step guides group-based decision-making for selection of the biomass, the agricultural supply chain (transformation operations and upstream/downstream processes) and technologies. Decisions may be taken manually or with the assistance of decision support tool, such as ELECTRE, or PROMETHEE. Mixed group-based decision-making methods, such as Delphi-SWOT (Tavana et al., 2012), will be implemented in the future development of this approach. The approach is illustrated with a case study in the next section: an analysis of four agricultural supply chains in the environmental area relating to rice straw valorisation processes in France.

3. Case study: valorisation of the rice straw supply chain

3.1. Pretreatment of lignocellulosic biomass

The lignocellulosic biomass has four main components: cellulose, hemicellulose, lignin, and phenolic acids. Cellulose and hemicellulose can be hydrolysed to generate glucose. Lignin and phenolic acids are responsible for the recalcitrance of cellulosic materials, the crystallinity of cellulose and the particular surface and porosity characteristics of matrix polymers. Biomass pretreatment is therefore essential, to decrease crystallinity, to increase the specific surface area and porosity and to extract the major constituents. Various pretreatment methods have been studied in detail over the last 30 years. Each of these pretreatment methods, whether mechanical, physical, chemical, physicochemical, biological or a mixture of various types, has advantages and disadvantages. Various factors have been used to compare the performance, efficiency or environmental impact of these pretreatment processes: environmental factors, energy consumption and energy efficiency, for example (Barakat et al., 2014; Chuetor et al., 2015; Zhu and Pan, 2010). Biomass pretreatment process studies used a cradle-to-gate approach (Jacquemin et al., 2012), extending from the pre-milling of the biomass to its enzymatic hydrolysis (Fig. 4).

The goal of this case study was to use the approach described above and its associated workflow to help researchers to choose a sustainable process for the valorisation of rice straw pretreatment. The preliminary results obtained are presented here. The LCA method was applied to the environmental domain for the *sustainability analysis*. Future studies will focus on the economic area, and will make use of the life cycle costing (LCC) method. This study was a cradle-to-gate. The system boundaries were set at the pre-milling and enzymatic hydrolysis steps. The transport of the biomass from the field to the firm was taken into account. The core hypotheses were: (i) a pre-pilot-scale process is studied (ii) rice straw is considered to be free agricultural waste with no environmental impact – all impacts of the rice crop are attributed to the part used for food (iii) the energy is French mixed electricity (v) the site of the study is France, and the field and the firm are a known distance apart. The functional unit chosen was “the production of 1 g of glucose”. Glucose yield was required for the expression of flows per functional unit. Four different processes for treating rice straw were studied. These processes consisted of combinations of the four transformation operations shown in Fig. 4:

(a) RSP1 (rice straw process 1), with one operation: pre-milling.
(b) RSP2, with two milling operations: pre-milling + ultrafine milling. Like the previous process, this is a mechanical process.
(c) RSP3, with three operations: pre-milling + physicochemical treatment + pressing and separation.
(d) RSP4, with the four operations in the sequence shown in Fig. 4: pre-milling + physicochemical treatment + pressing and separation + ultrafine milling.

The last step of the transformation process is enzymatic hydrolysis treatment, which was the same for all four systems studied. The only data from this hydrolysis used were glucose yield and the amount of buffer, which is dependent on biomass quality (more buffer required for lower biomass quality).

3.2. Results

The general workflow (Fig. 3) is illustrated for the case study in Appendix A. The first step of the workflow (*goal*) is specified in Section 3.1.

In the *data architecture* step, data from various big data sources were collected and extracted to complete the assessment and to create decision support for the researcher: public web, field data, corporate data, LCA databases and scientific databases (Appendix A). The scientific databases included articles published on the four rice straw processes. Fifteen articles were selected by biomass industrial engineering experts. The LCA database usedwas Ecoinvent, which contains background data and information about the uncertainties on these data. The field database contained all the flow data, for both input and output, for the operation of the process, together with information about the technologies used for individual operations. Data were extracted from these databases in the form of CSV files. These CSV files were then used for the second substep the *enrichment and*

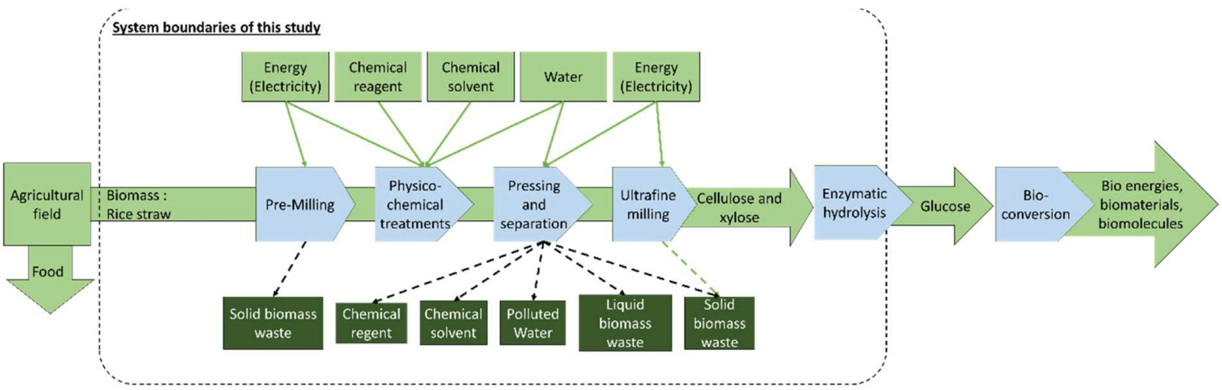


Fig. 4. Presentation of the by-products valorisation processes.

storage of all data, corresponding to add data. The milling unit operation was identified for each article, but the energy data were missing, hampering the *sustainability analysis* (third step). Internal empirical models created by rice straw pretreatment experts were available and were used to enrich the data.

The third substep, *data curation*, was performed manually by sustainability engineering experts. Two curations were performed with the 15 rice straw articles. The material balance was not checked in two of the 15 articles selected due to lack of information. All the data from these two articles were, therefore,

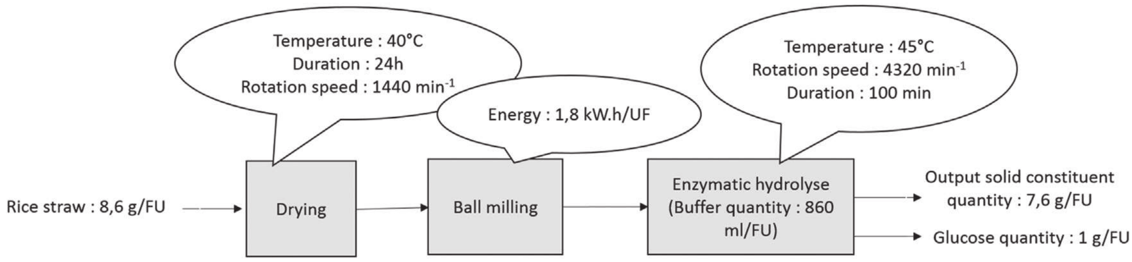


Fig. 5. Example of data visualisation from an "average experiment".

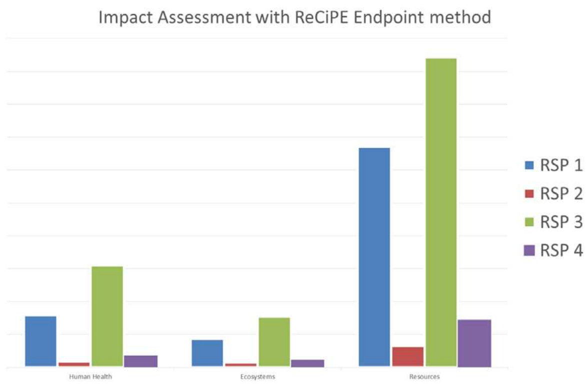


Fig. 6. Visualisation at endpoint level.

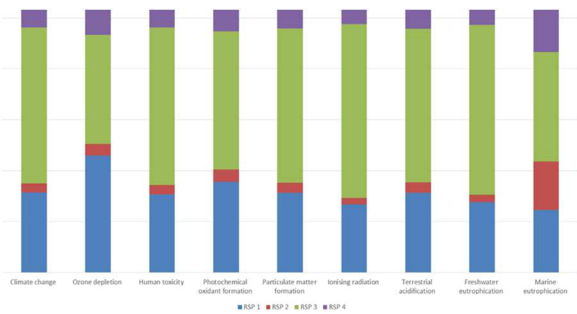


Fig. 7. Visualisation at midpoint method (1/2).

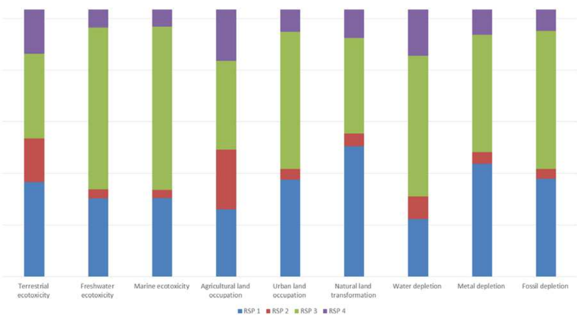


Fig. 8. Visualisation at midpoint method (2/2).

deleted. The availability of energy data for unit operations was also very patchy. We therefore decided to remove the energy data presented in some of the articles, to make it possible to upgrade all articles and avoid bias. This upgrade allowed to compare the different industrial path for the next steps (*sustainability analysis, sustainability visualisation* and *decision*).

After this curation, the fourth substep was *data analytics*. The 13 articles described 39 experiments. We decided to create one experiment from each article to facilitate subsequent calculations, analysis, visualisation 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 (Fig. 5 gives an example). These 13 "average experiments" were grouped together into four types of process: RSP1, RSP2, RSP3 and RSP4.The *data visualisation* substep generated the inventory tables used for the next step. At the end of the *data architecture* step life-cycle inventories for each process (RSP1, RSP2, RSP3 and RSP4) were generated.

The third step, *sustainability analysis,* uses the ReCiPe 2016 method to assess environmental impact and the Monte-Carlo method to calculate dispersion. The ReCiPe impact factor database (RIVM, 2018) and the Monte-Carlo method are available from the public web. The LCA database used was Ecoinvent (EcoInvent Life Cycle Inventory Database, 2017). The pedigree matrix was obtained from scientific articles and was completed for background data in Ecoinvent. We chose to use the LCA method and associated ISO standards. The data for the foreground system, resulting from step 2, and the background data, extracted from Ecoinvent, were mixed to calculate the environmental effects of each input and output. The ReCiPe method was then used to calculate 18 midpoint indicators and three endpoint indicators.

The fourth step, *sustainability visualisation*, provided a visualisation of these environmental indicators (Figs. 6–8). The different types of figure show the impact assessment at the midpoint or endpoint level. A standard representation of midpoint indicators is available in Figs. 7 and 8 presents a common representation of endpoint indicators. These interactive visualisations supported decision-making for the last step, *decision*, concerning the "best" pretreatment process for rice straw valorisation. The endpoint method uses three indicators: human health, ecosystems and resources. Results are expressed per functional unit, making it possible to compare different processes. RSP1 and RSP3 had the greatest impact on the three indicators considered. This impact can be explained by the pre-milling operation for RSP1. The production of 1 g of glucose with this pretreatment requires both a large amount of biomass and lots of buffer for the enzymatic hydrolysis. The use of such a large amount of buffer has a very high impact. In RSP3, the pressing and separation operation had the highest impact. Thus, two very different processes were found to have high endpoint indicators. With this visualisation, it was not easy to identify the process with the least environment impact, because the differences between RSP2 and RSP4 were minor.

The visualisation of midpoint indicators can be more relevant for decision-making (Figs. 7 and 8). RSP2 appeared to have a lower impact than RSP4 for all but three indicators: marine eutrophication, terrestrial ecotoxicity and agricultural land occupation. At this point, the decision-making group, could either make a decision directly or refer to additional studies comparing RSP2 and RSP4. However, biorefinery chains are innovative and few data relating to biorefineries have been published to date. Further studies are therefore required.

The team of data scientists and researchers in industrial and sustainability engineering applied the five steps to test the approach. Results were obtained with a newly created research tool coupling a well-known LCA application (Simapro®) and a Microsoft Excel-VBA application. This tool supports all the steps of the proposed approach except the decision step, with decisions taken by a decision group. This approach could be improved for the case study presented. Each new article about rice straw pretreatment will provide additional process data, thereby improving the sustainability analysis. Thermodynamic models are then required to complete the data from scientific publications, particularly for the calculation of energy consumption for the process operations. The case study could also be broadened by the inclusion of the economic area in future developments of this research tool.

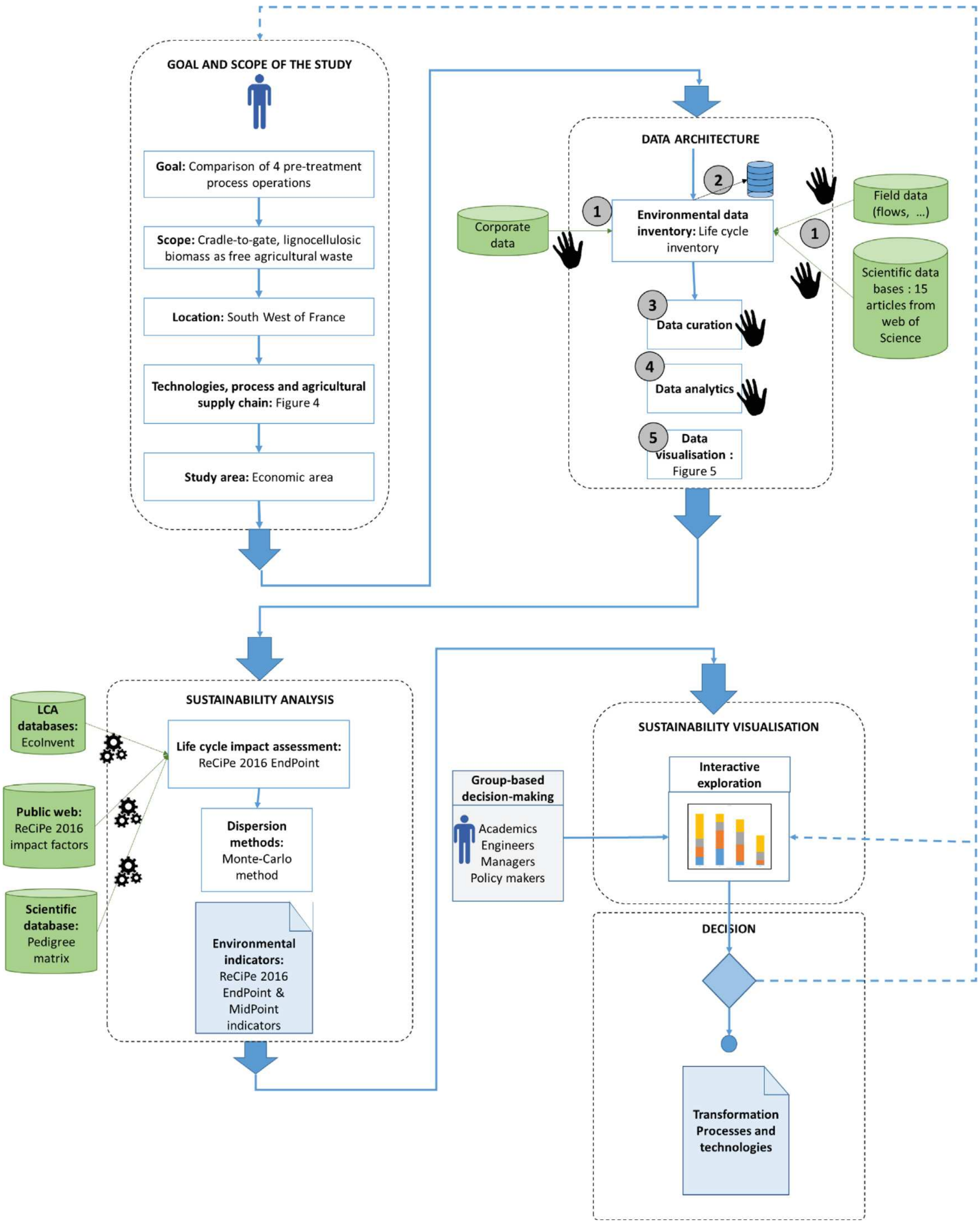
4. Conclusion and outlook

We have designed an approach integrating industry 4.0 into supply chain design to improve sustainability management for the valorisation of agricultural waste. This five-step approach combines methods and tools from big data and sustainability assessment. The goal, limits and hypotheses of the study are first specified. This makes it possible to determine which data are required for the second step: data architecture. Based on big data architecture, five substeps of the data architecture step have been defined: data extraction, storage, curation, analysis and visualisation. These five substeps provide all the data required for the next step: sustainability analysis. The sustainability visualisation step can then generate results through various dynamic and in-depth visualisation techniques. Finally, the decision step involves either a group-based decision-making process or a semi-automatic decision method. Three areas of sustainability — economic, environmental and social — can be assessed. In the case study presented, we applied this approach to the assessment of four pretreatment processes in the agro-food industry. This approach ruled out two processes as having too great an environmental impact. Additional studies would be required to enable the decision group to identify the "best" of the remaining processes. Various possibilities for improving this approach and case study could be explored: (i) addition of specific data sources, methods and visualisations for the economic and social areas, to improve sustainability data inventories and assessment methods; (ii) progress towards the automation of data extraction for step 2. This would make it possible to save time and to add new sources of data more easily; (iii) from our feedback with the Excel-VBA research tool, development of a complete ergonomic computing framework supporting the approach. This would encourage stakeholders to adopt this approach and would facilitate decision-making through the implementation of collaborative decision-making techniques, such as Delphi-SWOT; (iv) the design of models for calculating energies; (v) the generalisation of this principle and the development of a library of business and domain-specific models from agro-food process engineering. These models could be used to check and validate the data in the data architecture step. Controls could include, for example, an advanced material balance or energy analysis; (vi) the development of data dispersion propagation and automatic qualitative explanation systems for stakeholders. These advances will help to refine the method, to render it more general and more accurate.

Declaration of Competing Interest

None.

Appendix A. Rice Straw Valorisation Workflow



References

Babiceanu, R.F., Sekr, R., 2016. Big data and virtualization for manufacturing cyber-physical systems: a survey of the current status and future outlook. Comput. Ind. 81, 128–137.

Barakat, A., Chuetor, S., Pouau, X., Solhy, A., Rouau, X., 2014. Eco-friendly dry chemo-mechanical pre-treatment of lignocellulosic biomass: impact on energy and yield of the enzymatic hydrolysis. Acs Appl. Energy Mater. 113, 97–105.

Belaud, J.-P., Negny, S., Dupros, F., Michéa, D., Vautrin, B., 2014. Collaborative simulation and scientific big data analysis: illustration for sustainability in natural hazards management and chemical process engineering. Comput. Ind. 65, 521–535.

Belaud, J.-P., Adoue, C., Vialle, C., Sablayrolles, C., 2019. A circular economy and industrial ecology toolbox for developing an eco-industrial park: perspectives from French policy. Clean Technol. Environ. Policy 21, 967–985.

Bhinge, R., Srinivasa, A., Robinson, S., Dornfeld, D., 2015. Data-intensive life cycle assessment (DILCA) for deteriorating products. Procedia CIRP 29, 396–401.

Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M., 2014. How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective. Int. J. Mech. Ind. Sci. Eng. 8 (1), 37–44.

Chen, H., Chiang, R.H.L., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36 (4), 1165–1188.

Chen, C., Zhang, C.-Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf. Sci. 275, 314–347.

Chuetor, S., Luque, R., Solhy, A., Rouan, X., Barakat, A., 2015. Innovative combined dry fractionation technologies for rice straw valorization to biofuels. Green Chem. 17, 926–936.

Cooper, J., Noon, M., Jones, C., Kahn, E., Arbuckle, P., 2013. Big data in life cycle assessment. J. Ind. Ecol. 17 (6), 796–799.

De Mauro, A., Greco, M., Grimaldi, M., 2016. A formal definition of big data based on its essential features. Libr. Rev. 65, 122–135.

Doan, A., Halevy, A., Ives, Z., 2012. Principles of Data Integration. Morgan Kaufmann Publishers.

EcoInvent Life Cycle Inventory Database, 2017. EcoInvent Life Cycle Inventory Database. Available: http://www.ecoinvent.org/. (Accessed 24 May 2017).

Food SCP Round Table European Commission, 2012. Continuous Environmental Improvement - Final Report. .

Garnett, T., Appleby, M., Balmford, A., Bateman, I., Benton, T., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., Herrero, M., Hoffmann, I., Smith, P., Thornton, P., Toulmin, C., Vermeulen, S., Godfray, H., 2013. Sustainable intensification in agriculture: premises and policies. Science 341 (6141), 33–34.

Ghisellini, P., Cialani, C., Ulgoati, S., 2016. A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. J. Clean. Prod. 114, 11–32.

Hampton, S., Strasser, C., Tewksbury, J., Gram, W., Budden, A., Batcheller, A., Duke, C., Porter, J., 2013. Big data and the future of ecology. Ecol. Environ. 11 (3), 156–162.

Hardaker, J.B., 1997. Guidelines for the Integration of Sustainable Agriculture and Rural Development. 2. The Concept of SARD. FAO, Food and Agriculture Organization, Rome.

Horton, P., Koh, L., Shi Guang, V., 2016. An integrated theoretical framework to enhance resource efficiency, sustainability and human health in agri-food systems. J. Clean. Prod. 120, 164–169.

ISO, 2006. Environment Management - Life Cycle Assessment - Principles and Framework. 14040:2006. International Organization for Standardization, Geneva, Switzerland.

Jacquemin, L., Pontalier, P.-Y., Sablayrolles, C., 2012. Life cycle assessment (LCA) applied to the process industry: a review. Int. J. Life Cycle Assess. 17, 1028–1041.

Kagermann, H., Wahlster, W., Helbig, J., 2013. Recommendations for Implementing the Strategic Initiative Industry 4.0 - Final Report of the Industry 4.0 Working Group. Acatech, München.

Levin, O., Ketner, J., Krapohl, D., 2013. NIST Big Data Public Working Group: Reference Architecture Subgroup. Available: https://bigdatawg.nist.gouv/_uplloadfiles/NIST.SP.1500-6.pdf. (Accessed 13 September 2018).

Liao, Y., Lezoche, M., Panetto, H., Boudjilida, N., Rocha Loures, E., 2015. Semantic annotation for knowledge explication in a product lifecycle management context: a survey. Comput. Ind. 71, 24–34.

Mathews, J., Tan, H., 2011. Progress towards a circular economy: the drivers and inhibitors of eco-industrial initiative. J. Ind. Ecol. 15, 435–457.

Ming, X., Cai, H., Liang, S., 2015. Big data and industrial ecology. J. Ind. Ecol. 19 (2), 205–210.

Noy, N., 2004. Sematic integration: a survey of ontology-based approaches. SIGMOD Rec. 33 (4), 65–70.

Palmer, M., Bernhardt, E., Chornesky, E., Collins, S., Dobson, A., Duke, C., Gold, B., Jacobson, R., Kingsland, S., Kranz, R., Mappin, M., Martinez, M., Micheli, F., Morse, J., Pace, M., Pascual, M., Palumbi, S.S., Reichman, O., Townsend, A., Turner, M.G., 2005. Ecological science and sustainability for the 21st century. Front. Ecol. Environ. 3, 4–11.

Pavwelczyk, A., 2005. EU Policy and Legislation on recycling of organic wastes to agriculture. Int. Soc. Anim. Hyg. 1.

Pivoto, D., Waquil, P.D., Talamini, E., Spanhol Finocchio, C.P., Dalla Corte, V.F., de Vargas Mores, G., 2018. Scientific development of smart farming technologies and their application in Brazil. Inf. Process. Agric. 5, 21–32.

Raymond, R., 2012. Improving food systems for sustainable diets in a green economy. Greening the Economy with Agriculture, 2012 United Nations Conference on Sustainable Development: Governance for Greening the Economy with Agriculture .

Reichman, O., Jones, M.B., Schildhaueur, M., 2011. Challenges and opportunities of open data in ecology. Science 11, 703–705.

N. i. f. P. H. a. t. Environment, 2018. N. i. f. P. H. a. t. Environment. RIVM Available: https://www.rivm.nl/en/Topics/L/Life_Cycle_Assessment_LCA/Downloads/Documenten_ReCiPe_2016/. (Accessed 26 September 2018).

Snijders, C., Matzat, U., Reips, U.-D., 2012. "Big data": big gaps of knowledge. Int. J. Internet Sci. 7 (1), 1–5.

Tavana, M., Pirdashti, M., Kennedy, D.T., Belaud, J.-P., Behzadian, M., 2012. A hybrid Delphi-SWOT paradigm for oil and gas pipeline strategic planning in Caspian Sea basin. Energy Policy 40, 345–360.

West, P., Gerber, J., Engstrom, P., Mueller, N., Brauman, K., Carlson, K., Cassidy, E., Johnston, M., 2014. Leverage points for improving global food security and the environment. Science 345, 325–328.

Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.-J., 2017. Big data in smart farming - a review. J. Agric. Food Syst. Commun. Dev. 153, 69–80.

Zhang, Y., Luo, X., Buis, J., Sutherland, J., 2015. LCA-oriented semantic representation for the product life cycle. J. Clean. Prod. 86, 146–162.

Zhu, J., Pan, X., 2010. Woody biomass pre-treatment for cellulosic ethanol production: tehcnology and energy consumption evaluation. Bioresour. Technol. 101 (13), 4992–5002.