



Article

# Data Lake Governance: Towards a Systemic and Natural Ecosystem Analogy

Marzieh Derakhshannia <sup>1</sup>, Carmen Gervet <sup>2</sup>, Hicham Hajj-Hassan <sup>3,\*</sup>, Anne Laurent <sup>1</sup>  
and Arnaud Martin <sup>4</sup>

<sup>1</sup> LIRMM, Univ. Montpellier, CNRS, 34090 Montpellier, France; dm.derakhshannia@gmail.com (M.D.); anne.laurent@umontpellier.fr (A.L.)

<sup>2</sup> Espace Dev, Univ. Montpellier, IRD, Univ Guyane, Univ. Réunion, 34293 Montpellier, France; carmen.gervet@umontpellier.fr

<sup>3</sup> CNRS-L, Beirut P.O. Box 11-8281, Lebanon

<sup>4</sup> CEFE, Univ. Montpellier, CNRS, 34293 Montpellier, France; arnaud.martin@umontpellier.fr

\* Correspondence: hishamhh@cnrs.edu.lb

Received: 12 May 2020; Accepted: 29 June 2020; Published: 27 July 2020



**Abstract:** The realm of big data has brought new venues for knowledge acquisition, but also major challenges including data interoperability and effective management. The great volume of miscellaneous data renders the generation of new knowledge a complex data analysis process. Presently, big data technologies provide multiple solutions and tools towards the semantic analysis of heterogeneous data, including their accessibility and reusability. However, in addition to learning from data, we are faced with the issue of data storage and management in a cost-effective and reliable manner. This is the core topic of this paper. A data lake, inspired by the natural lake, is a centralized data repository that stores all kinds of data in any format and structure. This allows any type of data to be ingested into the data lake without any restriction or normalization. This could lead to a critical problem known as data swamp, which can contain invalid or incoherent data that adds no values for further knowledge acquisition. To deal with the potential avalanche of data, some legislation is required to turn such heterogeneous datasets into manageable data. In this article, we address this problem and propose some solutions concerning innovative methods, derived from a multidisciplinary science perspective to manage data lake. The proposed methods imitate the supply chain management and natural lake principles with an emphasis on the importance of the data life cycle, to implement responsible data governance for the data lake.

**Keywords:** data lakes; data governance; sustainability; supply chain management; natural lake; ecosystem

## 1. Introduction

With the realm of big data as a source of new knowledge extraction through data analysis and mining techniques, machine learning, correlation, and cluster analysis techniques, data heterogeneity and interoperability are common challenges. Ontologies and Findable, Accessible, Interoperable, and Reusable (FAIR) systems are presently able to handle these challenges effectively [1]. However, another challenge is rising, concerning the data volume and storage lifespan. In the past few decades, due to the vast amount of data being generated each second, data storage systems and analytical tools play a vital role in the big data ecosystem. They facilitate the processes of storing, manipulating, analyzing, and accessing structured and unstructured data (J. Dixon. Pentaho, Hadoop, and data lakes. <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>, 2010).

Among modern data storage systems and repositories, we are primarily interested here in data lakes, designed to store a large volume of data in any format and structure. The data lake is a

recent generation of storage systems conceived as data repositories to propose a flexible platform for data storage, access exploration, and analysis [2,3]. Because their existing features can handle data heterogeneity, they provide means to generate new knowledge and identify data patterns from large amounts of data, independently of their format and structure. According to Fang, a data lake is a cost-efficient data storage system enabled by the new generation of data management technologies to master big data problems and improve data analyzing process from ingestion, storage, exploration and exploitation of data in their native format to mining information and extract new knowledge from massive unstructured data [4]. A data lake uses a flat architecture to collect and store data, on a platform initially based on Apache Hadoop (Highly Available Object-Oriented Data Platform) which is a beneficial big data tool [5,6].

As mentioned above, a data lake operates as a central repository which loads data in no-schema approach. This means that the data is ingested into the data lake without predefined structure and the schema is defined only at the time of data usage and data querying. This approach is known as schema-on-read or “late bindings” and is the opposite of schema-on-write which is common in data warehouse [4,7]. In data lakes, the “extreme volume” of raw data is stored and processed at the lowest possible cost, unlike data warehouses that load large scale of “cleansed” data in a more costly manner [4]. According to Sawadogo and Darmont, a data lake could be viewed as a form of data warehouse that collects multiple structured data with minimum operational cost before extract-transform-Load (ETL) process or as a global storage system that contains a data warehouse for enhancing data life cycle monitoring with “cross-reference analyses” [7].

A successful data lake must satisfy properties concerning data handling and management such as: cost-effective and flexible ingestion, storage, processing, data access, and “applicable data governance policies” [8].

In general, data lakes contain heterogeneous and multi-modal data that renders its analysis complex, and sets requirements for rigorous processes to maintain and ensure data integrity from its storage to its exploitation. This will allow us to improve the data quality for data scientists and to decrease the cost of data storage and risk. Hence the concept of data governance has resurfaced to support the mastering of data management, to control data quality, and improve business intelligence in insurable manners [9]. Nevertheless, the life cycle of data that enters a data lake is seldom accounted for. There is a strong need to conceive, define, and implement data governance mechanisms, to handle proper data retention and minimize the risk of data swamps.

According to Madera and Laurent, “data governance is concerned with the data life cycle, quality and security of data” [2] in any storage system. Hence it is a fundamental issue in relation to the data lake authenticity. Data governance disciplines and strategies aim to prevent data lakes from becoming data swamps or maintaining poor-quality data [10]. These disciplines control or fix data quality dimensions, like: “accuracy”, “completeness”, “consistency”, “currency” and “uniqueness” to guarantee validity of data [11] and complement data management [12,13]. In the big data ecosystem, many governance mechanisms have been proposed to guarantee the veracity and accuracy of the data value. In particular, Abraham, Schneider, and Vom Brocke [14] distinguish three categories for data governance mechanisms which are frequently implemented for data management:

- Structural mechanism in references to the governance structures
- Procedural mechanism related to the policies for data management
- Relational mechanism concerned with stockholder communications

With reference to this specification, some researchers define some standards or guidelines to manage data in data repositories [15]. Others like Yebenes and Zorrilla [16] propose frameworks for big data management. Some researchers put the emphasis on communication agreements to deploy feasible data governance [17]. Since data access can be a strong competitive advantage for any organization and is shared to exchange information, Van den Broek and Van Veenstra [18] presented some regulations to govern and balance data contributions. Data governance plays an important

role in improving self-service business intelligence in the big data era [19]. Consequently, beneficial data governance guidelines could minimize the risk of poor data quality in data mining processes and improve its accuracy [10]. Data governance assessment improves the strategy frameworks for deploying successful data governance with respect to relevant focus areas [20]. A practical data governance framework, with a focus on data quality, increases confidence in exploration and exploitation of the data [13] and monitors *data quality efforts* in a sustainable manner [21]

All proposed mechanisms for data governance, concentrate on setting principles, roles, and structures to improve data quality and data lake security. The data life cycle management is one of the most important reasons for applying data governance in each data repository. However, the influence of data lifespan on proper data quality strategies for deleting or preserving data in the data lake has received little attention, even though mortality and life expectancy of data in the data lake is a serious issue when it comes to increasing the productivity of the data lake.

In this paper, we start from the assumption and claim that data governance implementation concerning the data life cycle, could influence the general purification of data repositories from useless data.

The concept of data lake is defined as a **system** with multiple components that are derived from **natural lake definition**. Hence, some data governance policies or regulation methodologies have been extracted from systematic approaches or natural mechanisms to preserve and destroy data throughout their life cycle. This viewpoint provides enormous capabilities to govern a data lake effectively. In this article, we propose two solutions that are respectively derived from drawing analogies with (1) nature ecosystems and (2) the concept of the supply chain to address data lakes and their governance issues. *Our approach is based on a comparison of the dynamics, life cycles, and operations within those two systems with those needed for data lakes. We show that such perspectives provide paradigms for optimizing data lake performance, and we describe some methods for sustainable data governance.*

**Nature ecosystem analogy.** Let us consider living organisms and particularly the DNA. The information is determined by the activity of the “reader”. The data which is not read is not used (the principle). The data is not systematically destroyed after a “not-being-read” period, but if such a period becomes long, then the data is weakened and may disappear. Also, if the data is frequently read, then it is consolidated and solidified even if this can have a penalizing effect, later on.

What can happen in such situations, is that the individual or the species can disappear. However, at the same time, chance can create new data and multiply it. These are the characteristics of living things that can generate new data automatically. This “natural” mechanism can be implemented on data governance in the data lake. Please note that the notions of “long” or “chance” would then clearly need to be instantiated and specified. **Systematic approach analogy.** The goal of a systematic approach is to identify the most efficient means to generate consistent and optimum results [22]. Such approaches, implemented in the supply chain domain, are another analogy we draw to address our objectives. For instance, Chen and Huang [22] use a systematic approach to recognize the interaction between supply chain’s members as system elements. To do so, they decompose the supply chain participants into sub-groups and sub-system elements and enhance the supply chain structure to represent a complex system that will improve coordination and integration among supply chain elements.

Indeed, the strategies and methodologies which are frequently used in supply chain management bring practical paradigms for promoting service quality and resolving customer affairs issues. If we consider a data lake as a supply chain and, consequently data as a product, we could define a set of hybrid policies for improving data quality and thus reach an optimal data lake state. For example, lean management strategies provide some approaches to minimize additional costs and eliminate wastes in a data lake just by defining the costly activities or non-valuable data [23]. Similarly, a strategy frequently used in the supply chain such as “agile management”, will improve the responsibility and flexibility of the data lake with regard to user requirements with high quality of service even in critical situations [23].

Those two frameworks can be viewed as effective paradigms for managing the data life cycle, and its governance to ensure its viability. In the following, we postulate that with the assumption that data lakes are comparable to natural lakes and to supply chains, the processes derived from nature and supply chain management can be extrapolated to data management in repositories such as data lakes. Based on this positioning, we present a general analogy and comparison between supply chain management, natural lakes, and data lakes and identify similar aspects and components. Then, based on those similarities we propose new methodologies to improve data lake's validity.

## 2. Our Approach and Contribution

Based on the definition of a "system", the ecosystem and supply chain are both considered intelligent systems that contain several components and are governed by specific rules and disciplines. A data lake is conceptually inspired by a natural lake. Consequently, all concepts that are frequently used in the data lake, originate from a natural lake ecosystem. From another perspective, supply chain management provides some appropriate concepts and processes that are also applicable to data lake management and data governance.

Our study is based on the position that a data lake as a system, has many common and comparable elements with supply chains and natural ecosystems. Dealing with diverse and heterogeneous data in data lakes—like products in supply chains or species in ecosystems—requires hybrid solutions and methods for data management which can be accurately determined. In line with our focus on the data life cycle, we put the emphasis on designing practical methods to preserve valid data in the data lake and remove invalid or obsolete ones from the data lake. For instance, it is logical that some data will be separated from the data lake, like a defective product in supply chain. The data can also be brought back to the data lake or kept after its usage, like a reusable or recyclable product in close-loop or reverse supply chain [24] or like the information in backward flow across the chain. This addresses which data is concerned.

In addition, a key challenge lies in the evaluation of data usage during their lifespan, because a data lake stores data that may be retrieved or queried in the future, rather than serving an immediate need [25].

We would assume that data acts like products in the supply chain or water in a natural lake. Hence, they have a probabilistic lifespan and may be valid and useful (i.e., have high value for exploration and exploitation) or invalid and obsolete (i.e., have no value and increase the risk of data swamp). Therefore, to avoid storing invalid data and managing the data life cycle, we tackle the challenge of the data life expectancy by drawing analogies with processes used in the supply chain and ecosystem to govern data lakes.

The questions are then:

- Which aspects of the data lake are comparable with natural lake and supply chain?
- Which strategies should be derived from nature and supply chain for data governance?
- How should these strategies be generalized to a data lake?

In this article, we contribute some first research positions, by:

- Providing comparisons between a data lake, an ecosystem, and a supply chain (Element by Element);
- Relying on supply chain management strategies for data governance (Systematic Manner);
- Imitating nature principals to manage the data life cycle (Natural Manner).

## 3. Comparing Data Lake, Ecosystem and Supply Chain

Each system consists of different components that work effectively together to achieve certain goals under deterministic or probabilistic restrictions and conditions. Furthermore, each system applies strategies to optimize different objective functions and improve overall performance. The performance

of this system is evaluated according to several criteria to examine how many optimal levels have been fulfilled.

As previously mentioned, an ecosystem and the supply chain inherently act as a system, and in many aspects are comparable with each other. Similarly, a data lake, as a centralized storage system, behaves in accordance with analogous systematic paradigms. Regarding this point, we have elaborated tables that compare supply chain, data lake, and ecosystem with each other, thus explaining the relationship we identified between a set of concepts. Tables 1 and 2 present a general comparison of the three systems.

**Table 1.** Analogy of Supply Chain, Ecosystem and Data Lake.

Module	Supply Chain	Natural Lake (Species and Ecosystems)	Data Lake
Members/ Levels	Supplier Manufacturer Distributor Retailer Customer	Ecosystem Components (Animals, Plants, Microorganisms) Biological processes (Breed, Birth, Growth and Death) Ecological processes (Eat and be Eaten)	Ingestion stage Storage stage Processing stage Access stage Final user
Products	Commodity (Forward Flow) Information (Backward Flow)	Biodiversity (Species diversity) Ecological complexity (More Species More complex) Biomass	Data
Management Strategies	Lean SCM Agile SCM Postponement SCM Speculation SCM Green Supply Chain	Species evolution (Mutation, Recombination, Drift, Selection) Competition Parasitism (Negative association) Mutualism (Positive association) Predation	Metadata management Data management Data Governance
Objective Functions	Cost minimization Sales maximization Profit maximization Lead time minimization	At species level: Maximize reproduction and survive (Fitness) At ecosystem level: Maximize resilience	Cost minimization Fill rate maximization Response time minimization

Following the structure of Tables 1 and 2, we develop the different analogies further.

### 3.1. Supply Chain and Data Lake

Formally, a supply chain is a corporation of different entities such as manufacturers, suppliers, distributors, and retailers that cooperate to provide specific products or services for consumers [26]. To create a **profitable supply chain**, all members of the chain should be vertically integrated with all parties being coordinated across the optimal goal of the chain [27]. One of the major considerations in supply chain management is the integration of all members towards a global goal, and the improvement of the product flow and information across the chain. According to Simchi-Levi, Kaminsky & Simchi-Levi, Delfmann & Albers and Harland [28–30], supply chain management includes managerial techniques and processes to integrate all members of the chain, from suppliers to retailers, to minimize whole system expenditures, improve chain profit and increase service levels satisfaction. The first step in supply chain management is to define the **objective functions** of the chain that optimize the **decision variables** which are characterized by the supply chain manager.

Typically, supply chain objective functions intend to minimize expenditures [31,32], wastes, maintenance and storage cost, inventory cost, lead time and customer service time [33], and to maximize profit, coverage demand, and service levels [32]. The fundamental goal of supply chain management is to add value and provide a clear competitive advantage to enhance chain productivity

and efficiency. Meanwhile, to design, manage and evaluate an integrated supply chain, some major modules need to be accurately defined [34] :

- Chain members and their responsibilities (components or participants)
- Product
- Management strategies
- Objective function
- Decision variables
- Constraints
- Risks
- Qualitative performance measurement

As Tables 1 and 2 show all these modules which have been characterized for the supply chain could also be defined for a data lake if we consider it as an integrated system with certain components and stages.

**Table 2.** Problem Specifications of Supply Chain, Ecosystem and Data Lake

Module	Supply Chain	Natural Lake (Species and Ecosystems)	Data Lake
Decision Variables	Location Allocation Order quantity Volume Service sequence Number of levels(stages) Number of facilities	Homeostasis	Service sequence Number of sets Number of data scientists Number of reservoirs
Constraints	Budget Number of warehouses Number of active facilities Capacity Service compliance Lead time	Lake emergence Perturbations Global changes	Number of data set Capacity Data gravity Data governance principles Service compliance
Risks	Risk of losing the customer Risk of defective product Risk of information failure Risk of quality failure Risk of overstock products Risk of high delay time Risk of machine failure	Strong perturbations	Data failure Machine failure Security Unreliable data Access control Risk of machine failure
Qualitative Performance Measurement	Customer satisfaction Transaction satisfaction Flexibility Information integration Lead time minimization Supplier performance Manufacture performance Transportation performance	Species richness Ecosystems functions Resilience	Data quality Data lake flexibility Better data acquisition Quick access to raw data Data preservation Agility

With respect to the first module, each member (level) of the supply chain is responsible for the specific task of enhancing the value of the whole chain. Suppliers must provide the best raw material, manufacturers must produce high-quality products, distributors are responsible for logistics

management and retailers improve the service levels for the final customers. The result of the member's collaboration is the optimal and integrated supply chain with high customer satisfaction. Likewise, according to LaPlante & Sharma [35], four major functions are described for data lakes, from data entry to its preparation for the final user (typically data scientists). These functions are divided into four principal stages: ingestion, storage, processing, and access stages which organize data in levels. Ingestion management, controls of data sources (where data come from), data storage (where data are stored), and the data arrival time (when data arrive). Ravat & Zhao [36] also proposed a "data lake functional architecture", which is structured with four main zones: Raw data zone, process zone, access zone, and governance zone. Regarding these proposed architectures for data lake, we can consider data lake as a supply chain that collects, generates, transfers, and delivers data from several resources to the final users.

The second module is **product**. The major products in the supply chain are commodities in forward flow and information in backward flow. However, in the data lake, the products are data that can be considered to be commodities or information in the supply chain. Considering this point, the main products of the data lake are the data with an appropriate management plan from their ingestion level to the information extraction.

In the third module of Table 1, the main purpose of this comparison is **management strategies**, that are defined as a set of improvement plans and patterns which are used for enhancing system performance and providing the specific principles and objectives to reach the goals [37,38]. Consequently, all other modules in the supply chain, like: parameters, objective function, decision variables, and constraints of the chain, will be determined based on a relevant strategy. For example, green strategies are applied to the supply chain, to minimize the environmental cost and maximize the green-conscious customer satisfaction [39,40]. Similarly, for data lake management, some strategies, like data governance and metadata management, are frequently used to accomplish definitive goals and increase data quality.

As mentioned in Table 1, **objective function** is the important module that impacts on subsequent decisions in the supply chain management [34]. Accordingly, cost minimization and profit maximization are two important objectives that the whole supply chain seeks to reach. Similar objectives are also common in data lakes. The goal of maximizing or minimizing the objective function is to obtain the optimal value for the decision variables with respect to the constraints of the problem. The type of these decision variables differs in the supply chain and in the data lake, but they have the same meaning.

As we can see from Table 2, the number of facilities or warehouses are the critical **decision variables** in each supply chain, and making decisions about them is a strategic and long-term decision [34]. Similarly, in data lakes, the optimal number of repositories or sets must be estimated accurately. **Risk management** plays a vital role in system management and is determined according to the internal and external conditions of each system [41]. In general, the risk of machine failure or defective product in the supply chain and risk of data swamp and unreliable data in the data lake are the most prominent risks. Finally, **performance evaluation** is essential for system development. Therefore, some evaluation standards are specified according to the characteristics of the systems [27,34].

From both tables showcasing our comparison between the supply chain management and data lake, it is obvious that both systems have been generated for similar purposes which are:

- Improving integration between members and information
- Reducing waste [3,42,43]
- Achieving the agility, flexibility, and sustainability [23,44]
- Increasing service levels

Therefore, there are very similar points between the supply chain and the data lake. Thus, it seems logical that supply chain tools and strategies can be efficient to enhance the data lake performance and productivity. In this article, we propose to use one of the most successful assessment methods,

presented in Section 4, used to monitor the environmental performance of the supply chain. We intend to use it to implement data governance according to the life and death of data in data lakes.

### 3.2. Ecosystem and Data Lake

In this second analogy, we are considering the lake as an ecosystem filled with numerous living species. These species are the members of our system. They have different functions. For example, some species eat others. All species have a common feature: they reproduce and survive. However, the system is more than the sum of its members, and that is what we will detail.

The ecosystem is seen as an autonomous system whose regulations are not necessarily aimed at the survival of all species, but to guarantee the homeostasis and resilience of the system. Homeostasis is permitted by sets of regulations [45]. Biologists consider that resilience is linked to the complexity of the system, the number of species, and the number of internal regulations [46,47]. Thus, biologists consider that the more complex the system is, the healthier it is.

In our comparison, the essential point is homeostasis (decision variable in the table), and we will consider resilience as an underlying property of the system. On the scale of a living organism, homeostasis operates through a complex set of regulations according to a simple principle of three functions: a receiver, a control center, and an effector. In the case of an ecosystem, the mechanisms are more complex [48,49], and the ecologists are currently just able to analyze precisely the relations between homeostasis and resilience and their role in the stability of the system. For our study, we will retain that the ecosystem has internal regulatory functions that maximize its survival and good health. These functions are not determined by a system supervisor, but by the system itself.

The results of the comparison sections demonstrate that a data lake could be defined as an integrated system based on supply chain terms and ecosystem regulations in which all related members act coherently. In the next sections, concerning the table interpretation, we distinguish the methods of data governance in the data lake in two manners: supply chain-based method or systematic manner and ecosystem-based method or natural manner, to suggest two multidisciplinary solutions for managing the data life cycle in data repositories such as data lakes.

### 3.3. Examples

We provide here some detailed examples to illustrate our contribution and to point out to some further research we will carry out that will rely on the tables we exhibit as a result.

#### 3.3.1. Members/Levels

The supply chain is a connection of multiple dependent or independent members or levels that contribute to each other with a common goal of adding value to a product or service from sources to destinations [27]. For example, in a three-level supply chain, the three principal members are: manufacture, distributor, and retailer [28]. In data lakes, each stage acts as a member in a supply chain to provide (APIs, data and service endpoints), transport (IP addressing, ...), store (HDFS file system, ...) and make data accessible for the final users [35]. In biological systems, the levels are those of life, from DNA sequences to cells, bacteria, species, ... which are called ecosystem components. Therefore, these members, whether they belong to a supply chain, a data lake, or an ecosystem, are responsible for product quality and service levels.

#### 3.3.2. Products

A broad range of products exists in the supply chain network, for instance, seasonal products like clothing, alimentary products like canned food or industrial products like machines, which are logistically managed with specific standards and fixed lifespan [28].

In biological systems and natural lakes, products can be DNA sequences, species, or biomass which are reproduced and preserved by certain mechanisms in nature.



Similarly, in data lakes, data is a targeted product which could be sensor data, web log data, financial data, human or machine-generated data that must be stored and managed, with a given logistics.

In all these three systems, products can be considered at different levels of granularity, as components or complex systems.

### 3.3.3. Management Strategies

Each supply chain regarding its objective, type of product, structure, and market demand, is managed with a specific strategy [28]. For example, seasonal or perishable products like clothing or fresh food respectively, with a very short life cycle, do need concrete planning to increase the product sale during their lifespan hence agile strategy could be an effective solution [23].

On the other hand for ecological products in the environmental supply chain, some specifications like a recyclable product or not, and some other considerations along the logistical process lead the green supply chain to derive numerous solution strategies [39,40].

In the ecosystem, the main strategies for species evolution are mutation, recombination, selection, and drift.

In analogy, data in data lake concerning their structure and utility, need to follow certain regulations, relative to entering a data lake and its possible usages. The goal is to ensure the quality of the data mining process, by deriving suitable data governance as a management strategy responsible for guaranteeing data quality.

### 3.3.4. Objective Functions

Objective functions are defined and aligned with opted management strategies for designing supply chain networks. For example in the supply chain with seasonal products, the objective function could be service level maximization or response time minimization; or in the green supply chain, we would define the minimization of CO<sub>2</sub> emission or total cost minimization [50].

On the other hand, the maximization of species reproduction and resilience of the ecosystem are major considerations of the ecological system.

Thus, the main objective function is related to minimizing poor data quality and maximizing the customer's usage rate.

### 3.3.5. Decision Variables

Regarding the definition of decision variables, a set of decision variables is commonly considered in supply chain optimization models. For instance, in the seasonal product supply chain, a decision variable can be the amount ordered, and in the green supply chain, the degree of environmental protection [50].

Similarly, in the ecosystem, homeostasis is a key decision variable, for which we seek the optimal value.

In analogy, important decision variables in data lake management are defined as the total amount of satisfied demands or the number of users that are permitted to access the data lake.

### 3.3.6. Constraints

Constraints distinguish the scope of the optimization model. For example, the lead time is a critical constraint in the supply chain with seasonal products, and environmental level constraints are essential to reason about green supply chains [50].

In the ecosystem, critical constraints like global changes that are induced by some drivers like CO<sub>2</sub> enrichment and *biotic invasions*, could restrain optimal interactions between species [51].

In a data lake, the laws of gravity and data governance principles are the most important limitations that describe the problem boundaries.

### 3.3.7. Risks

For seasonal products, the risk of losing the customer is definitely of high impact due to the short lifespan of products. In green supply chains, the risk of the data with destructive effects is significant [31,41].

Some remarkable risks like hydrologic perturbations which are derived from climate change could have a serious impact on ecological systems [52].

Thus, in a data lake, storing unreliable data or data failure are major risks.

### 3.3.8. Qualitative Performance Measurement

Qualitative performance measurement is essential for evaluating any system efficiency, and to examine actual gaps between the existing and the desired system [27]. For example, customer satisfaction or rate of Flexibility are characterized as qualitative performance measurements for seasonal product supply chain, and the degree of adaptation of the chain to environmental standards for green supply chain [40].

Resilience and *optimal ecosystem functionality* are important quantitative qualifications which are determined by diversity measures like *response diversity* [53].

Similarly, agility, data quality, and data lake flexibility could be determined as fundamental qualitative measurements to evaluate data lake performance.

## 4. Data Governance in Supply Chain

Supply chain management is related to strategies and rules that integrate all upstream and downstream relationships across the chain to generate high levels of value for direct and indirect participants [54]. Recently, environmental responsibility has received increasing attention as an inseparable element for every supply chain to remove or reduce the non-biological products that have a dangerous impact on the environment and natural cycle. Based on these requirements, several strategies and disciplines, like green supply chain management [39,40] and environmental supply chain management [55], are defined.

Environmental supply chain management (ESCM) is related to the sustainable strategies that use life cycle assessments (LCA) from raw materials to final customers and the reverse flow of products (recycle or disposal) [55]. The LCA is an instrument based on an environmental consideration that monitors and restricts the destructive environmental effects of a product's entire life cycle in the supply chain with specific standards [54,56,57]. Based on such instruments, other completed assessment codes and procedures, like: PLCA (product life cycle assessment) [58], SLCA (social life cycle assessment) [59] and LCSA (life cycle sustainability assessment) [58,60], are proposed by different organizations [61].

The purpose of all proposed assessment codes is to regulate the whole procedure throughout the supply chain, in order to eliminate or minimize the harmful impacts on the environment. Each one of these standards assesses a specific aspect of the product life cycle, such as social or cost aspects, for instance. The monitoring of a product's life cycle with such protocols improves the internal performance and productivity of the supply chain and consequently expands ecological and social care with cost-effective products.

Due to the data life cycle in the data lake, such assessment codes could serve as infrastructure for data governance legislation. Based on this cognition, data assessment is implemented from data collection to data interpretation, and all poor-quality or useless data, which have no value for data lake or data mining, will be limited or prevented from entering data lake. From our point of view, by regulating specific codes for data life cycle assessments (DLCA), data lake will be purified from life to death of data under strict disciplines.

The International Organization for Standardization (<https://www.iso.org>) defines ISO 14040 as "Code of Practice" for life cycle assessment which includes four major phases in LCA study [62]:

- *The goal and scope definition phase*
- *The inventory analysis phase*
- *The impact assessment phase*
- *The interpretation phase [63]*

These phases could be extended for data in the data lake to implement data life cycle assessments (DLCA). According to the goal and scope definition phase, we should determine which data with which qualifications is targeted, in order to address target users, system boundary, data category, and targets for data quality [63]. In the inventory analysis phase, all information about the quality of input and output data is collected and validated under the life cycle assessment study of data. Then in the impact phase, all information about the effects of various data quality on the data lake, based on impact categories and life cycle inventory results, is evaluated. Finally, the impacts of different data quality on the data lake are interpreted with respect to some features like “validity”, “sensitivity” and “consistency” of data. The final results are concluded or reported in the interpretation phase. Consequently, this approach ensures data quality for the data’s lifespan with accurate assessment protocol [64,65].

## 5. Data Governance in Natural Ecosystem

For most biologists, the basic building block is the gene. It is the unit that contains living information. Richard Dawkins [66] explains that living things are made up of genes that reproduce through envelopes, organisms, which are simple avatars of genes. One may wonder why there are so many different life forms. We share identical genes with many species (97% homology with great apes, like chimpanzees or gorillas). Certain fundamental genes, like for example the one that codes for hemoglobin, for instance, is almost identical in very many species. However, ecosystems are very diverse, and they appear to us to be relatively stable structures where information seems to be constantly organized, distributed, and redistributed.

Considering that even before the appearance of the first cells, self-replicating molecules have existed and living things reproduce with a prolixity far above the level of acceptance of the system. There are therefore regulations that are carried out by the mechanism of natural selection (only the ablest survive). Natural selection is the constraint of living things. During reproduction, sexuality allows the mixing of genes and introduces a factor of chance (in addition to other phenomena such as, for example, mutations). Thus, the two forces which frame living beings are chance and necessity (constraint).

Chance does not produce information. It only produces complexity. Necessity is what produces information [67]. Take moving animals for example, elephants cross a forest to seek a resource. This action will be repeated over the generations. The first animal makes its way “at random”, the second also, and so on. Soon enough, paths will exist and will be taken by the following elephants because it is less expensive in terms of energy to follow a path than to create a new one. Then there will be a selection of the most practical paths (to bypass natural obstacles, for example). In the end, there remains a reduced path network that forms an optimum choice for the shortest and least costly path. This network results from the effect of necessity (go to the least costly). The combined action of chance and necessity, conditions not only information as it is observed, but also its evolution [67]. If once again, we take the paths created by the elephants, we can consider that at any moment, the chance can engage the evolution in a new way, while the necessity will force the new way to remain functional.

## 6. Conclusions

A data lake, as a complex storage system, needs a variety of methods to govern heterogeneous data accurately and in a timely manner. In this article, we have proposed some multidisciplinary approaches, which are **natural manner** and **systematic manner**, for data governance in data lake and argued that supply chain strategies and natural principals could be the effective sources of inspiration

for data governance in order to assess the *life cycle of data from the moment they enter the data lake until they are destroyed*.

First, we provided a comparison table to indicate that the data lake acts as a system and has some aspects similar to the supply chain and ecosystem, both referred to as a complex system. Therefore, we considered data in data lakes, like products in the supply chain or species in nature, to draw similarities and identify proper strategies for data governance.

Then, we proposed two different methods based on systematic methods and natural behaviors to suggest a new perspective of data governance in the big data environment. Our methodology and comparative analysis showed that life cycle assessment codes as a systematic approach and revival of the laws of nature were ideal multidisciplinary approaches to implement sustainable data management with respect to life and death of data.

Proposed methods are derived from different disciplines and our contribution for comparing and aligning concepts impacts all data lake components and processes, from data collection to data exploitation. For these reasons, there are some limitations to examine their concrete exploitation for data lakes within one single work. We rather consider that this work opens many research avenues to consider every comparison and every data lake component one by one.

Therefore, with regard to our conclusion, we propose some future case studies for implementing our work in the real world and evaluating the obtained results.

One study will consider, for instance, data lake performance optimization. For this reason, we will use the design of Supply Chain Network strategies to define a mathematical model that maximizes the service level of the data lake; since supply chain management optimizes the profit. For this reason, a proper strategy like agile strategy will be opted and objective function(s) which maximizes the service, decision variable(s) such as the amount of satisfied demand, and constraint(s) such as the capacity or budget will be determined accordingly. Choosing a suitable strategy and designing the components of a mathematical model is an important challenge that must be carefully considered.

We will implement our proposed framework using real data lake software. We will consider and evaluate several aspects of data collection, data storage, and data processing in data lakes. As mentioned in Section 4, for implementing this approach for data life cycle assessment, four major phases should be determined. In future work, we will develop these four phases in a data lake to deploy a practical perspective on data governance. However, it is essential to distinguish qualitative or quantitative measurements for describing valuable and destructive data in order to monitor good- or poor-quality data in the data lake.

Another work with regard to the principal objectives of this article could be inspired by the lean strategy in the supply chain to minimize the total cost of poor-quality data in data lakes. For this purpose, we aim to define the **cost** in the objective function, reducing the impact of all data that have no value for data lake or increase the risk of the data swamp. Decision variable(s) and constraint(s) for this mathematical model will be determined, respectively.

Finally, based on our analogy table, we will use biological models to recommend and manage relevant and promising data localization in the file systems, data crossings, etc., as DNA and biological materials do in nature.

**Author Contributions:** Conceptualization, M.D., C.G., A.L. and A.M.; Funding acquisition, H.H.-H.; Methodology, M.D., C.G., A.L. and A.M.; Project administration, A.L.; Validation, C.G., A.L. and A.M.; Writing—original draft, M.D., A.L. and A.M.; Writing—review and editing, C.G., H.H.-H. and A.L. All authors have contributed to the concepts, content, writing, revisions and proofreading of this article. All authors have read and agreed to the published version of the manuscript

**Funding:** This work was partly funded by the PHC CEDRE project Number 42415YJ co-funded by the French Ministry of European and Foreign Affairs (MEAE), French Ministry of Higher Education, Research and Innovation (MESRI) and Lebanese Ministry of Education and Higher Education (MEHE).

**Acknowledgments:** This work was partially supported by PHC CEDRE 42415YJ, French Ministry of European and Foreign Affairs (MEAE), French Ministry of Higher Education, Research and Innovation (MESRI) and Lebanese Ministry of Education and Higher Education (MEHE). This work is also a contribution to the HUT Project (HUMAN at home Project).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Nat. Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
2. Madera, C.; Laurent, A. The next Information Architecture Evolution: The Data Lake Wave. In Proceedings of the 8th International Conference on Management of Digital EcoSystems, Biarritz, France, 1–4 November 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 174–180. [CrossRef]
3. Russom, P. Data lakes: Purposes, practices, patterns, and platforms. In *TDWI White Paper*; Talend: Redwood City, CA, USA, 2017.
4. Fang, H.L. Managing data lakes in big data era: What's a data lake and why has it become popular in data management ecosystem. In Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), Shenyang, China, 8–12 June 2015; pp. 820–824.
5. Khine, P.; Wang, Z. Data Lake: A New Ideology in Big Data Era. Available online: [https://www.itm-conferences.org/articles/itmconf/pdf/2018/02/itmconf\\_wcsn2018\\_03025.pdf](https://www.itm-conferences.org/articles/itmconf/pdf/2018/02/itmconf_wcsn2018_03025.pdf) (accessed on 2 July 2020).
6. White, T. *Hadoop: The Definitive Guide*, 4th ed.; O'Reilly: Sebastopol, CA, USA, 2015.
7. Sawadogo, P.N.; Darmont, J. On Data Lake Architectures and Metadata Management. *J. Intell. Inf. Syst.* **2020**. to appear. [CrossRef]
8. Gorelik, A. *The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science*; O'Reilly Media: Sebastopol, CA, USA, 2019.
9. Ladley, J. *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*; ITPro Collection; Elsevier Science: Amsterdam, The Netherlands, 2012.
10. Paschalidi, C. Data Governance: A Conceptual Framework in Order to Prevent Your Data Lake from Becoming a Data Swamp. Available online: <https://itu.diva-portal.org/smash/record.jsf?pid=diva2%3A1019917&dswid=2135> (accessed on 2 July 2020).
11. Loshin, D. Chapter 5—Data Governance for Big Data Analytics: Considerations for Data Policies and Processes. In *Big Data Analytics*; Loshin, D., Ed.; Morgan Kaufmann: Boston, MA, USA, 2013; pp. 39–48. [CrossRef]
12. Wende, K. A Model for Data Governance—Organising Accountabilities for Data Quality Management. p. 80. Available online: <https://www.alexandria.unisg.ch/publications/67284> (accessed on 2 July 2020).
13. Al-Ruithe, M.; Benkhelifa, E.; Hameed, K. Data Governance Taxonomy: Cloud versus Non-Cloud. *Sustainability* **2018**, *10*, 95. [CrossRef]
14. Abraham, R.; Schneider, J.; Vom Brocke, J. Data governance: A conceptual framework, structured review, and research agenda. *Int. J. Inf. Manag.* **2019**, *49*, 424–438. [CrossRef]
15. Aisyah, M.; Ruldeviyani, Y. Designing data governance structure based on data management body of knowledge (DMBOK) Framework: A case study on Indonesia deposit insurance corporation (IDIC). In Proceedings of the 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS 2018), Yogyakarta, Indonesia, 27–28 October 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 307–312. [CrossRef]
16. Yebenes, J.; Zorrilla, M. Towards a Data Governance Framework for Third Generation Platforms. *Procedia Comput. Sci.* **2019**, *151*, 614–621. [CrossRef]
17. Allen, C.; Des Jardins, T.R.; Heider, A.; Lyman, K.A.; McWilliams, L.; Rein, A.L.; Turske, S.A. Data Governance and Data Sharing Agreements for Community-Wide Health Information Exchange: Lessons from the Beacon Communities. *J. Electron. Health Data Methods* **2014**, *2*, 1057. [CrossRef]
18. Van den Broek, T.; Van Veenstra, A.F. Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. *Technol. Forecast. Soc. Chang.* **2018**, *129*, 330–338. [CrossRef]

19. Riggins, F.J.; Klamm, B.K. Data governance case at KrauseMcMahon LLP in an era of self-service BI and Big Data. *J. Account. Educ.* **2017**, *38*, 23–36. [[CrossRef](#)]
20. Panian, Z. Some practical experiences in data governance. *World Acad. Sci. Eng. Technol.* **2010**, *38*, 150–157.
21. Thomas, G. *The DGI Data Governance Framework*; Data Gov. Institute: Orlando, FL, USA, 2006; Volume 20.
22. Chen, S.J.; Huang, E. A systematic approach for supply chain improvement using design structure matrix. *J. Intell. Manuf.* **2007**, *18*, 285–299. [[CrossRef](#)]
23. Ciccullo, F.; Pero, M.; Caridi, M.; Gosling, J.; Purvis, L. Integrating the environmental and social sustainability pillars into the lean and agile supply chain management paradigms: A literature review and future research directions. *J. Clean. Prod.* **2018**, *172*, 2336–2350. [[CrossRef](#)]
24. Guide, V.D.R.; Van Wassenhove, L.N. OR FORUM—The Evolution of Closed-Loop Supply Chain Research. *Oper. Res.* **2009**, *57*, 10–18. [[CrossRef](#)]
25. Derakhshannia, M.; Gervet, C.; Hajj-Hassan, H.; Laurent, A.; Martin, A. Life and Death of Data in Data Lakes: Preserving Data Usability and Responsible Governance. In Proceedings of the Internet Science—6th International Conference (INSCI 2019), Perpignan, France, 2–5 December 2019; Yacoubi, S.E., Bagnoli, F., Pacini, G., Eds.; 2019; Volume 11938, pp. 302–309. [[CrossRef](#)]
26. Heaven, T.; Chow, G. Logistics strategies for North America in the Global logistics and distribution planning. In *the Global Logistics and Distribution Planning*; Waters, D., Ed.; Kogan Page Limited: London, UK, 1999.
27. Janvier-james, A.M. A New Introduction to Supply Chains and Supply Chain Management: Definitions and Theories Perspective. *Int. Bus. Res.* **2012**, *5*, 194–207. [[CrossRef](#)]
28. Simchi-levi, D.; Kaminsky, P.; Simchi-Levi, E. *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*; McGraw-Hill/Irwin: New York, NY, USA, 2003.
29. Delfmann, W.; Albers, S. *Supply Chain Management in the Global Context*; Working Paper 102; Cologne Publisher: Cologne, Germany, 2000.
30. Harland, C. Supply Chain Management: Perceptions of Requirements and Performance in European Automotive Aftermarket Supply Chains. Ph.D. Thesis, University of Warwick, Coventry, UK, 1994.
31. Azaron, A.; Brown, K.; Tarim, S.; Modarres, M. A multi-objective stochastic programming approach for supply chain design considering risk. *Int. J. Prod. Econ.* **2008**, *116*, 129–138. [[CrossRef](#)]
32. Nekooghadirli, N.; Tavakkoli-Moghaddam, R.; Ghezavati, V.; Javanmard, S. Solving a new bi-objective location-routing-inventory problem in a distribution network by meta-heuristics. *Comput. Ind. Eng.* **2014**, *76*, 204–221. [[CrossRef](#)]
33. Sawik, T. On the fair optimization of cost and customer service level in a supply chain under disruption risks. *Omega* **2015**, *53*, 58–66. [[CrossRef](#)]
34. Beamon, B.M. Supply chain design and analysis:: Models and methods. *Int. J. Prod. Econ.* **1998**, *55*, 281–294. [[CrossRef](#)]
35. LaPlante, A.; Sharma, B. *Architecting Data Lakes*; O'Reilly Media: Sebastopol, CA, USA, 2016.
36. Ravat, F.; Zhao, Y. Data Lakes: Trends and Perspectives. In Proceedings of the International Conference on Database and Expert Systems Applications (DEXA 2019), Linz, Austria, 26–29 August 2019; Volume 1, pp. 304–313.
37. Nickols, F. Strategy, Strategic Management, Strategic Planning and Strategic Thinking. *Manag. J.* **2008**, *1*, 4–7.
38. Ambe, I.; Badenhorst-Weiss, J. Framework for choosing supply chain strategies. *Afr. J. Bus. Manag.* **2011**, *5*. [[CrossRef](#)]
39. Tseng, M.L.; Islam, M.S.; Karia, N.; Fauzi, F.A.; Afrin, S. A literature review on green supply chain management: Trends and future challenges. *Resour. Conserv. Recycl.* **2019**, *141*, 145–162. [[CrossRef](#)]
40. Srivastava, S.K. Green supply-chain management: A state-of-the-art literature review. *Int. J. Manag. Rev.* **2007**, *9*, 53–80. [[CrossRef](#)]
41. Ritchie, R.; Brindley, C. Supply chain risk management and performance: A Guiding framework for future development. *Int. J. Oper. Prod. Manag.* **2007**, *27*, 303–322. [[CrossRef](#)]
42. Christopher, M. Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service (Second Edition). *Int. J. Logist. Res. Appl.* **1999**, *2*, 103–104. [[CrossRef](#)]
43. Miloslavskaya, N.; Tolstoy, A. Application of Big Data, Fast Data and Data Lake Concepts to Information Security Issues. In Proceedings of the 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Vienna, Austria, 22–24 August 2016. [[CrossRef](#)]

44. Sundaram, D.; Vidhya, M. Data Lakes-A New Data Repository For Big Data Analytics Workloads. *Int. J. Adv. Comput. Res.* **2016**, *7*.
45. Berntson, G.G.; Cacioppo, J.T.; Bosch, J.A. From Homeostasis to Allodynamic Regulation. In *Handbook of Psychophysiology*, 4th ed.; Cambridge Handbooks in Psychology, Cambridge University Press: Cambridge, UK, 2016; pp. 401–426. [[CrossRef](#)]
46. Karp, D.S.; Ziv, G.; Zook, J.; Ehrlich, P.R.; Daily, G.C. Resilience and stability in bird guilds across tropical countryside. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 21134–21139. [[CrossRef](#)]
47. Solow, A.; Duplisea, D. Testing for Compensation in a Multi-species Community. *Ecosystems* **2007**, *10*, 1034–1038. [[CrossRef](#)]
48. Morgan Ernest, S.K.; Brown, J.H. Homeostasis and Compensation: The Role of Species and Resources in Ecosystem Stability. *Ecology* **2001**, *82*, 2118–2132. [[CrossRef](#)]
49. Cottingham, K.; Brown, B.; Lennon, J. Biodiversity may regulate the temporal variability of ecological systems. *Ecol. Lett.* **2001**, *4*, 72–85. [[CrossRef](#)]
50. Wang, F.; Lai, X.; Shi, N. A multi-objective optimization for green supply chain network design. *Decis. Support Syst.* **2011**, *51*, 262–269. [[CrossRef](#)]
51. Tylianakis, J.M.; Didham, R.K.; Bascompte, J.; Wardle, D.A. Global change and species interactions in terrestrial ecosystems. *Ecol. Lett.* **2008**, *11*, 1351–1363. [[CrossRef](#)]
52. Kane, D.L. The Impact of Hydrologic Perturbations on Arctic Ecosystems Induced by Climate Change. In *Global Change and Arctic Terrestrial Ecosystems*; Oechel, W.C., Callaghan, T.V., Gilmanov, T.G., Holten, J.I., Maxwell, B., Molau, U., Sveinbjörnsson, B., Eds.; Springer New York: New York, NY, 1997; pp. 63–81. [[CrossRef](#)]
53. Mori, A.S.; Furukawa, T.; Sasaki, T. Response diversity determines the resilience of ecosystems to environmental change. *Biol. Rev.* **2013**, *88*, 349–364. [[CrossRef](#)] [[PubMed](#)]
54. Hagelaar, G.J.L.F.; Van der Vorst, J.G.A.J. Environmental Supply Chain Management: Using Life Cycle Assessment To Structure supply chains. *Int. Food Agribus. Manag. Rev.* **2001**, *4*, 399–412. [[CrossRef](#)]
55. Zsidisin, G.A.; Siferd, S.P. Environmental purchasing: A framework for theory development. *Eur. J. Purch. Supply Manag.* **2001**, *7*, 61–73. [[CrossRef](#)]
56. Sonnemann, G.; Castells, F.; Schuhmacher, M.; Hauschild, M. Integrated Life-Cycle and Risk Assessment for Industrial Processes. *Int. J. Life Cycle Assess.* **2004**, *9*, 206–207. [[CrossRef](#)]
57. Nwe, E.S.; Adhitya, A.; Halim, I.; Srinivasan, R. Green Supply Chain Design and Operation by Integrating LCA and Dynamic Simulation. In *20th European Symposium on Computer Aided Process Engineering*; Pierucci, S., Ferraris, G.B., Eds.; Elsevier: Amsterdam, The Netherlands, 2010; Volume 28, pp. 109–114. [[CrossRef](#)]
58. He, B.; Luo, T.; Huang, S. Product sustainability assessment for product life cycle. *J. Clean. Prod.* **2019**, *206*, 238–250. [[CrossRef](#)]
59. Toniolo, S.; Tosato, R.C.; Gambaro, F.; Ren, J. Chapter 3—Life cycle thinking tools: Life cycle assessment, life cycle costing and social life cycle assessment. In *Life Cycle Sustainability Assessment for Decision-Making*; Ren, J., Toniolo, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 39–56. [[CrossRef](#)]
60. Zanni, S.; Awere, E.; Bonoli, A. Chapter 4—Life cycle sustainability assessment: An ongoing journey. In *Life Cycle Sustainability Assessment for Decision-Making*; Ren, J., Toniolo, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 57–93. [[CrossRef](#)]
61. Mesaric, J.; Šebalj, D.; Franjkovic, J. Supply Chains In The Context of Life Cycle Assessment and Sustainability. *Bus. Logist. Mod. Manag.* **2016**, *16*, 53–70.
62. International Organization for Standardization (ISO). *14040: 1997—Environmental Management—Life Cycle Assessment—Principles and Framework*; International Organization for Standardization (ISO): Geneva, Switzerland, 2003.
63. Lee, K.; Inaba, A.; Sanöppu, K.S.T.; Asia-Pacific Economic Cooperation; Committee on Trade and Investment. *Life Cycle Assessment: Best Practices of ISO 14040 Series*; APEC Publication, Center for Ecodesign and LCA(CEL), Ajou University: Suwon-si, Korea, 2004.
64. Rebitzer, G.; Ekvall, T.; Frischknecht, R.; Hunkeler, D.; Norris, G.; Rydberg, T.; Schmidt, W.; Suh, S.; Weidema, B.; Pennington, D. Life cycle assessment part 1: Framework, goal and scope definition, inventory analysis, and applications. *Environ. Int.* **2004**, *30*, 701–720. [[CrossRef](#)]
65. Zhang, B.; Su, S.; Zhu, Y.; Li, X. An LCA-based environmental impact assessment model for regulatory planning. *Environ. Impact Assess. Rev.* **2020**, *83*, 106406. [[CrossRef](#)]

66. Dawkins, R. *The Selfish Gene*; Oxford University Press: New York, NY, USA, 2006.
67. Dessalles, J.; Gaucherel, C.; Gouyon, P. *Le fil de la vie—La face Immatérielle du Vivant*; Odile Jacob: Paris, France, 2016. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



© 2020. This work is licensed under <http://creativecommons.org/licenses/by/3.0/> (the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.