


SURVEY PAPER

Open Access



# Big data monetization throughout Big Data Value Chain: a comprehensive review

Abou Zakaria Faroukhi<sup>1</sup>, Imane El Alaoui<sup>2</sup>, Youssef Gahi<sup>1\*</sup>  and Aouatif Amine<sup>1</sup>

\*Correspondence:

gahi.youssef@uit.ac.ma

<sup>1</sup> Laboratoire Génie des Systèmes, Ecole Nationale des Sciences Appliquées de Kenitra, University of Ibn Tofail, Kenitra, Morocco  
Full list of author information is available at the end of the article

## Abstract

Value Chain has been considered as a key model for managing efficiently value creation processes within organizations. However, with the digitization of the end-to-end processes which began to adopt data as a main source of value, traditional value chain models have become outdated. For this, researchers have developed new value chain models, called Data Value Chains, to carry out data driven organizations. Thereafter, new data value chains called Big Data Value chain have emerged with the emergence of Big Data in order to face new data-related challenges such as high volume, velocity, and variety. These Big Data Value Chains describe the data flow within organizations which rely on Big Data to extract valuable insights. It is a set of ordered steps using Big Data Analytics tools and mainly built for going from data generation to knowledge creation. The advances in Big Data and Big Data Value Chain, using clear processes for aggregation and exploitation of data, have given rise to what is called data monetization. Data monetization concept consists of using data from an organization to generate profit. It may be selling the data directly for cash, or relying on that data to create value indirectly. It is important to mention that the concept of monetizing data is not as new as it looks, but with the era of Big Data and Big Data Value Chain it is becoming attractive. The aim of this paper is to provide a comprehensive review of value creation, data value, and Big Data value chains with their different steps. This literature has led us to construct an end-to-end exhaustive BDVC that regroup most of the addressed phases. Furthermore, we present a possible evolution of that generic BDVC to support Big Data Monetization. For this, we discuss different approaches that enable data monetization throughout data value chains. Finally, we highlight the need to adopt specific data monetization models to suit big data specificities.

**Keywords:** Value Chain, Big Data, Big Data Value Chain, Big Data Management, Big Data Monetization

## Introduction

Value creation is the rational goal of any organization. However, in a changing world, in terms of data, it is very difficult to find strategies to sustainably generate new value and improve operations. In this respect, generating value may mean revisiting some deeply-held conceptions of different processes. For this, each organization has to ensure an efficient and updated management of value creation processes for both inter-organization and partners. Value chain (VC) is one of the key solutions to such dilemmas.

Initiated by Michael Porter [1, 2], the VC consists of analyzing all activities as well as their interactions, identifies the sources of a potential competitive advantage, and organizes processes that contribute directly or indirectly to creating added value for the organization. That means, value chain analysis involves breaking down the business into strategically important activities to understand their impact [3]. However, adapting processes has become more and more complex especially when facing new challenges such as activities' digitalization. Indeed, digitizing processes within organizations is more efficient and bring out an often-neglected source of value, namely information. Because of that, the traditional value chain concept has become inappropriate for intangible assets and has to adopt a data-centric approach instead of a product-centric or process-centric one. Therefore, Data Value Chain (DVC) has become more and more adopted by organizations that mainly rely on data to extract useful insight and correlations.

Data Value Chain is a mechanism that defines a set of repeatable processes to extract data's value step by step throughout its entire lifecycle from raw data to veritable insights. DVC consists of four main discrete steps:

- Data generation: Capture and record data;
- Data collection: Collect, validate and store data;
- Data analysis: Process and analyze data to generate new potential insights;
- Data exchange: Expose the data outputs to use, whether internally or externally with partners.

It should be noted that data management has changed a lot over the years. New storage and processing technologies have emerged, changing not only the amount of data we are able to store, but also the way we store and manage that data. This evolution was made possible by the advent of Big Data.

Big Data is a logical consequence of the importance that digital has taken in our lives where the data is multiplying at an unprecedented rate. Produced by our computers, mobile phones, payment tools, but also by the multiple sensors that now equip cars, roads or our houses, this data, in its raw form either structured or not, is transmitted to thousands of data centers that store, analyze and crosscheck it continually. Big Data is often defined by seven criteria, commonly referred to as 7Vs: Volume, Velocity, Variety, Veracity, Value, Variability, and Visualization. Furthermore, it brings together a set of advanced tools that support each V even in real time. But the real revolution does not stop at mere analysis or storage capabilities. Big Data tools are now able to predict behaviors or events, and even react accordingly [4]. This revolution has prompted organizations to rethink the way to create value by adopting a derivative of DVC which is Big Data Value Chain (BDVC). This latter is well aligned with Big Data ecosystem [5] and allows to extract hidden reliable values while relying on the strengths of Big Data. A Big Data system is often complex to design. It provides functions and interfaces to cope with the different phases of the digital data lifecycle, from creation to destruction. Therefore, it is very important to adopt an approach to break down a typical Big Data system into several phases to extract value. This decomposition is the BDVC. It generally consists of up to five distinct phases:

- Data acquisition: Data acquisition refers to the process of obtaining raw data;
- Data pre-processing: Data pre-processing involves a validation process, cleaning, reduction and data integration to prepare storage;
- Data storage: Data storage includes not only storage but also the management of large-scale datasets;
- Data analysis: Data analysis uses analytical methods or tools to model, inspect and mine data to extract value;
- Data visualization: Data visualization is a method of assisting data analysis. It is a meaningful representation of complex data to show hidden patterns;

Big Data-driven organizations, that relies on BDVC approaches, generate higher benefits than traditional organizations since, once data is generated, it can be mined multiple times and for different needs [6]. This data can be exploited by others in its different forms or built on and exploited many times over. Thus, an increasing number of organizations are claiming to monetize their data as exchangeable or salable services. For this, enterprises can implement an effective strategy for data to be monetized in the raw or service form.

Our main contributions in this paper are the following:

- We propose a generic and exhaustive BDVC that support data from generation to exposition phase.
- We propose some of the interesting research that tackle data monetization.
- We propose a possible integration of data monetization throughout BDVC.

In this purpose, we adopt a systematic literature review methodology. A systematic review provides a method of evidence-based practice. It aims to evaluate, interpret and synthesize researches by adopting a scientific, replicable and transparent process. This approach allows to locate relevant researches and to guarantee the quality and veracity of articles finally selected. The systematic literature review gives valuable insights by providing a summary of the available evidence. The guidelines proscribed by D. Tranfiel [7] are followed in conducting this study. The stages respected in this research are: Planning the review, Conducting a review as well as Reporting and dissemination.

In this survey paper, we present the different steps of both VC and DVC concepts by showing their advantages and drawbacks as well as how they are a good basis for emergent concepts such Big Data. Also, we highlight some BDVC models and discuss how they can be adopted for data monetization strategies.

The rest of this paper is organized as follows; “[Data Value Chain models](#)” section presents some VC and DVC models. In “[Big Data characteristics](#)” section, we provide an overview about Big Data and its different characteristics. In “[The Big Data Value Chain approach](#)” section, we propose and discuss a generic BDVC, able through its various phases to monetize the data intrinsically. “[Towards Big Data monetization](#)” section describes the advent of Big Data Monetization. Moreover, we show some pathways to setup data monetization by using some specific BDVC models. Finally, “[Conclusion](#)” section concludes the paper and provides some research outlooks.

### **Data Value Chain models**

The DVC, from the digitization of everyday life to the implementation of Internet of Things (IoT), is today a key element for the sustainable use of information systems which are more and more ubiquitous. It is not just about data analysis, but it is also a set of interrelated resources and processes that creates value based primarily on available records. However, the fact of splitting the activities into several processes to extract value was not introduced specially to manage data-based processes. The principle of value chains has been introduced well before to help companies extract value.

In the next Subsections, we present the most known value chain models and how they have evolved to accompany data trends in data-centric organizations.

### **Value Chains**

Value Chain is a set of interlinked resources and processes that begins with the acquisition of raw materials and extends to the delivery of valuable products. The notion of VC was developed by Porter in the 1980s [1, 2, 8]. It is an analytical tool that implements how to break down the business activities and identify the interaction between them. The purpose of this decomposition is to distinguish between main activities and support activities that are responsible for creating value. Products or services move successively through all activities in the chain and, for each activity, products or services acquire value. In this initial model of VC, Marketing and sales operations are considered as core activities, while human resources management, technology development and infrastructure are classified as secondary activities. That said, Porter's model considers the value generated by information exchange and neglects internal value-creating activities [9]. Several models have been then developed based on this concept to extend the idea of extracting value by using several processes. In what follows, we highlight some most known extended VC concepts:

### ***Knowledge Value Chain***

Knowledge is seen as a key success factor and the basis of competitive advantage [10]. It is seen as the most important strategic resource. The value embedded in the products or services is mainly due to the development of resources derived from organizational knowledge. In fact, the ability of an organization to produce can be seen as the integration and application of specialized knowledge collectively carried by individuals in the organization. Therefore, the notion of value is no longer oriented by the customer, as in the porter's VC, but by the incorporation of knowledge in the products or services in the productive process of the firm. Knowledge Management (KM) aims to manage this resource integration in the organization's processes [10, 11]. The KM is interested in knowledge as a strategic resource that optimizes the functioning of the productive processes of the firm. To help the success of KM, it is therefore useful to analyze the chain of knowledge integration in order to identify the different fundamental stages of enrichment of this cognitive resource. The set of these processes is the Knowledge Value Chain (KVC). Following the considerable development of KM in the 2000s, the concept of KVC appeared and was strongly debated. A KVC is defined as a set of knowledge management processes. It is thus a KM framework organizing basic knowledge management

processes. The main phases forming the KVC are: knowledge acquisition, knowledge storage, knowledge dissemination, knowledge application and knowledge sharing [10, 12]. KVC is often used to determine the viability and success of the organization in a competitive environment.

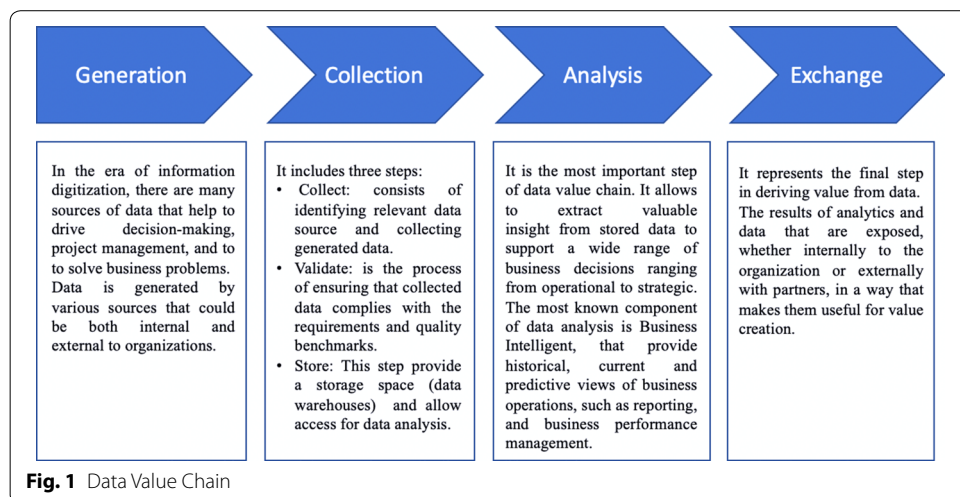
**Value grid**

Traditional value chains are well aligned with landline organizations, but opportunities today come in many different and unexpected ways. This argues for considering value creation as multidirectional rather than linear, which is called value grid. The grid approach allows organizations to go beyond traditional linear approaches and landlines industries by considering new processes for creating value. In a value-grid approach, there are a variety of possible pathways to create value. They can be vertical, horizontal, or even diagonal [13]. Unlike traditional linear VC, value grid allows decision makers to strategize and coordinate their operations. This model offers several strategies such as modifying information access, managing risk, integrating value, creating new value propositions and exploiting value chains across tiers [13, 14].

All VC models, presented above, operate in different contexts and present several ways to optimize processes in order to maximize profits and generate value from products and services. However, it is worth noting that with the advent of new technologies and digitization, products and services take on a non-tangible dimension, giving birth to another form of good, namely data. Organizations have been more and more oriented towards data processing to extract the most valuable insights and benefits. However, traditional value chain concept, that often-neglected data, has become inappropriate for intangible assets. So, new VC concept called DVC has emerged to fit data specifications.

**Data Value Chains**

The Data Value Chain is a model that sees data as a raw material as well as a major resource in the business. DVC describes a series of processes needed to generate step by



step value and useful insights. It consists of four main phases, described in Fig. 1, namely Data Generation, Data Collection, Data Analysis and Data Exchange.

There are also other data value models in the literature that are inspired from the above steps and provide enhanced approaches. In what follows, we present Linked data value chain and value chain network which are deemed to be good extensions of the initial data value chain concept:

#### ***Linked Data Value Chain (LDVC)***

In 2006, the term Linked Data was defined as a set of best practices for exposing, sharing and connecting structured data on the web using URIs [15] and RDF [16, 17] data model. The basic idea of linked data is to apply the global architecture of the World Wide Web to share structured data on a global scale. Then, these practices became known as linked data principles which are:

- The use of URIs to name the data;
- Using HTTP URIs;
- Provide useful information using RDF when the URI is visited; Include links to other URIs.

LDVC is a data value chain extension model based on concepts defining participating entities, their linked data roles as well as their processed data types [18]. It allows to collect data from many data sources such as raw data providers and offer that data to multiple application providers. Thus, LDVC allows to identify and categorize potential pitfalls and also helps to increase the value of production by forming a global value.

#### ***Data Value Chain network***

It refers to different actors that co-create and share data value through star network model, built around linked nodes under network-based business models [19]. Thus, value creation processes, based on data, have become more agile, dynamic and non-linear, with a multidimensional and traversal context, evolving in an environment that exchange information with other ecosystems [20, 21].

It should be recalled that DVC models consider data as a main resource for creating value. This allows many organizations to change the way they take decisions by adopting a data centric vision with specific data management strategies. However, data management has changed a lot over the years. In fact, new data characteristics have emerged, requiring not only new technologies and infrastructures to process data, but also the way to manage and create value from that data. Therefore, DVC had to follow this data revolution called Big Data in order to get the most from the daily generated data. In the next two Sections, we describe the world Big Data and its different characteristics, then we present the BDVC, the most suitable extension of DVC for Big Data era.

**Table 1** Commonly used approach in related work

Characteristic	Description
Volume	It refers to the size of data collected from different types of sources. Here, data is no longer represented by a few terabytes, but many petabytes. The bigger it is, the bigger is the insight we can mine
Velocity	It refers to the speed at which data is generated, stored and processed
Variety	Is one of the most challenges of Big Data. It concerns data produced from various sources, both internal and external. Data can be in different format; structured, semi-structured or unstructured
Veracity	It concerns anomalies and uncertainties in data, due to inconsistencies and incompleteness. The process of precluding bad data is therefore important to extract reliable insights
Variability	It is different from variety; it refers to data whose meaning is constantly changing
Visualization	Is the process of illustrating relationships within large amounts of complex data in a readable manner
Value	It refers to the information and insights that data provides

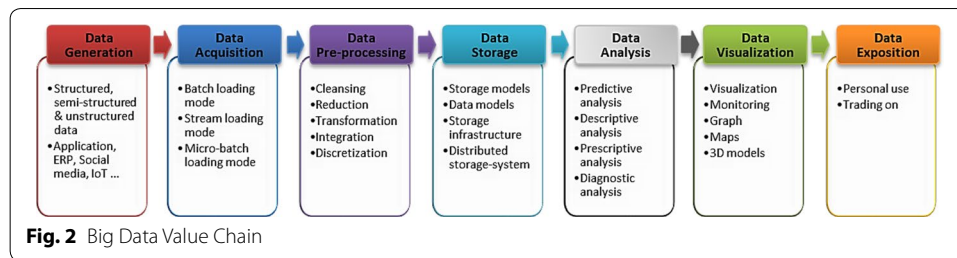
### Big Data characteristics

From 2010, the emergence of Big Data has brought enormous opportunities to the world of industry. Big Data is a logical consequence of digitalization which transforms various aspects of daily activities into digital records in order to provide an exceptional opportunity to mine valuable insights. However, generated data is no longer formed by only structured records, but it is characterized by any storage that has large volumes, high velocity, valid, varied and valuable content such as cookies, social activity, IoT records, videos, NoSQL Databases and gathered personal information. This kind of storage allows industries to extract valuable knowledge at the right time and even in real time [22–31].

Big Data characteristics have been originally defined by the 3Vs criteria (Volume, Variety and Veracity), then this vision has been extended over years and now often described using 7 criteria [32, 33] that we highlight in the Table 1.

It is becoming a crucial way for organizations to create value and realize their objectives by adapting their operating mode to deal with daily generated data. However, processing and analyzing such huge amount of heterogeneous records is not possible by using traditional tools and conventional methods, since it requires massive parallel processing techniques. Thus, it was mandatory to introduce novel techniques that are well aligned with new standards. These techniques and tools are called Big Data analytics. Big Data analytics consists of examining raw data, often in large volume, in order to extract information that is comprehensible to humans and difficult to observe by direct analysis. Big Data analytics propose a set of tools to deal with the data in its various states. These tools could be categorized into three main families: Storage, Processing and Visualization. Each of these families could be detailed by using appropriate tools according to the need and the level of required details.

We recall, Big Data revolution has prompted organizations to rethink the way they create value by adopting an extension of DVC named BDVC. This DVC extension is well aligned with the ecosystem Big Data and allows to extract hidden connections while relying on Big Data Analytics.



### The Big Data Value Chain approach

A Big Data system is a system that helps an organization to consider Big Data characteristics and enables to act on data to gain tangible benefits. It provides functions to cope with the different phases of the digital data lifecycle, from creation to destruction. Therefore, it is very important to adopt an approach to break down a typical process into several phases to extract value. This decomposition is named BDVC.

In the next subsection, we present a generic, global and exhaustive BDVC model that covers all required steps to deal with big data from generation to exposition. It is important to mention that those steps were separately addressed by other contributions discussed in the second Subsection.

### The proposed Big Data Value Chain model

The BDVC can be divided into seven steps, named respectively as shown in Fig. 2: Data Generation, Data Acquisition, Data Pre-Processing, Data Storage, Data Analysis, Data Visualization and Data Exposition. Besides, BDVC relies on various methods and techniques that we present throughout the Big Data Value Chain phases in the following.

#### Data generation

It is the first step of BDVC, it refers to the way in which data is generated. The actual source of that data could be internal or external. Besides, these data can be created passively or actively by humans, systems or sensors, and, can follow heterogeneous formats such as structured, semi-structured or unstructured [34, 35]. Once data generated, it will eventually power other systems through acquisition processes.

#### Data acquisition

It is the process of obtaining raw data. Data can be acquired in mainly three ways: batch mode, stream mode or micro-batch mode.

- The batch mode is an efficient mechanism for processing huge data over a period of time. This mode is recommended when the data is already stored in another traditional system, when the application does not require real-time processing or when the processing time has no influence on the result.
- The stream mode is the second method used to feed the storage system. It allows collecting and aggregating data from different sources in real time.



- The micro-batch mode is the intersection of the batch and the streaming modes. It divides streams into micro-batch. Consequently, the data is obtained within almost real time.

At the acquisition stage the data must be transferred from its capture point to a storage infrastructure using reliable networking infrastructure such as Content Distributed Network [32, 34, 36, 37]. Before storing that raw data, however, a pre-processing stage is highly required.

#### ***Data pre-processing***

Raw data which is collected from multiple sources may be completely noisy and redundant. Thus, it could clearly impact the analysis phase by producing meaningless results. Moreover, it may impact the data quality of raw data that is shared with other entities. Hence, data preprocessing appears to be of utmost importance for both analysis and exposition phases. The pre-processing phase consists of performing cleaning, qualifying, and improving data reliability. For this, multiple methods and techniques are available. Next, we present some of them:

- **Cleaning:** It consists of removing tuples with missing values by completing manually the data using a global constant, the mean value or the most probable value for the attribute. However, this method is not efficient when the percentage of missing values is important [37–40].
- **Reduction:** It consists of assuming that data follows a specific model, then estimate and store only the parameters of that model [37, 41]. It includes several methods, such as Big Data reduction [42], data noise reduction [43], data compression and decompression [44].
- **Transformation:** It converts the raw data to an appropriate format for further analysis by going through a series of specific steps [45].
- **Integration:** is the process of combining the data views in order to build a single view of data distributed over different sites [46]. However, it requires complex transformations and more computational capability before starting further phases [47].
- **Discretization:** It consists of dividing the range of possible values into sub-intervals [39, 45].

After preprocessing the raw data, data will be sent to a storage infrastructure, which is usually either a Datalake or an ordinary data center.

#### ***Data storage***

It consists of storing a huge amount of data that is collected from multiple sources. It is important to mention that the storage system highly impact the scalability and the performance of a BDVC. Thus, this step must provide reliable storage space and allow powerful access for data analysis. Generally, systems that provide Big Data storage operates around four pillars [32, 34, 38, 48–51]:

- **Storage models:** Is the mainstay of any system relying on Big Data. As any storage system could support three possible storage modes, Block-based, File-Based or Object-based, Big Data architecture relies on a block-based mode called HDFS.
- **Data models:** As mentioned above, traditional databases cannot meet the Big Data challenges (storing and processing huge amounts of unstructured data). For this, distributed storage and NoSQL databases are often adopted to manipulate persistent and heterogeneous databases. Distributed NoSQL databases [51] are grouped into four families: Key-value, Column-oriented, Graph-oriented and Document-oriented databases.
- **Storage infrastructure:** The hardware storage infrastructure for Big Data systems can be seen from different aspects such as storage devices, storage network infrastructure and storage virtualization.
- **Distributed processing infrastructure:** In computing, distributed systems consist of sharing data and tasks (calculation and processing) over several interconnected nodes [50]. Whereas, the user perceives the system as a single unit.

It is worth noting that all the above aspects allow facilitating data quality and data queries in favor of the analysis phase.

#### ***Data analysis***

It aims to analyze pre-processed stored data in order to find correlations, identify patterns and create actionable insights. There are mainly four categories through which Big Data analysis could be designed and conducted: prescriptive, predictive, diagnostic and descriptive [32, 36, 52–56]. In next, we describe each of these categories:

- **Descriptive analysis:** It provides insight into the past and identify what has happened.
- **Predictive analysis:** It aims to analysis scenarios of what might happen. The extracted insight is usually a predictive forecast. The predictive analytics brings together many data analysis technologies and other statistical techniques. Regression analysis is widely used in this area. It predicts the related values of multiple variables based on the confirmation or denial of a particular statement. Therefore, predictive analytics aim to recognize patterns in the data for the probability of a project.
- **Prescriptive analysis:** This kind of analysis reveals what decisions should be considered. This is the most valuable kind of analysis and usually results in rules and recommendations for next steps.
- **Diagnostic analysis:** It allows to identify the causes leading to the achievement of a performance by looking at past.

After this step, data are ready to be either visualized through meaningful dashboards or shared in its processed state.

#### ***Data visualization***

It allows to communicate insights drawn from data in a compelling and comprehensible fashion. In fact, there is an increasing desire to communicate information in a universal

way with an artistic visual representation such as graph, maps and 3D models. This would facilitate understanding and allow decision makers to explore datasets in order to identify correlations and hidden patterns [36, 57–59].

### **Data exposition**

It is the last phase in our proposed BDVC. It consists of sharing and exposing data, insights and knowledge created through out each phase. It is shared either internally or with partners, and sometimes with everyone as open data [60–64]. Data Exposition is often a part of a data strategy with two possible usage scenarios:

- Personal use: The organization does not really share the insights, but uses them to enhance and refine its own activities;
- Trading on: The organization uses the generated insights to sell specific services, such as targeted advertising.

It should be noted that each phase in BDVC uses specific Big Data analytics tools that are well aligned with that phase, either for storage, pre-processing, analysis or even exposition. It is also important to note that the nature and context of the project exhibit some unique characteristics and thus, require some adaptations in BDVC. That is, the order and the number of BDVC phases may differ depending on the use cases. To address these challenges, researchers have shown an increasing interest in developing different BDVC models.

At the best of our knowledge, no papers have been presented so far which propose an exhaustive BDVC from generation to exposition. In fact, several BDVC models have been proposed in the literature to manage and create value from Big Data as discussed in the following Section. Nevertheless, most of these studies propose basic model (basing on acquisition, storage and analysis) and completely ignore indispensable phases in big data such as exposition and monetization. Next, we highlight some of the important existing contributions on BDVC.

### **Big Data Value Chain models in the literature**

In what follows, we review some interesting models in which researchers have shown the best ways to deal with Big Data Value Chains context.

In business context, researchers in [65] has defined BDVC as a framework to manage data exhaustively, from capture to decision making, and to support a variety of stakeholders and also their technologies. The authors have aimed to manage and coordinate data across service continuity, streamlining data management activities and establishment of a portfolio management approach. This approach allows to achieve maximum combined value and to inform decisions in order to improve the organizations' performance. For this, they have proposed a model that consists of several phases namely: Data Discovery, Data Integration and Data Exploitation. Rajpurohit has proposed in [40] some Big Data processes and core components for business managers. The aim of these processes is to fill the gap between potential and value which is often due to the mismatch

between requirement and implementation. Besides, the critical errors across various phases, that have a severe impact on the quality of end-results, have been highlighted.

The authors in [5] have defined BDVC as a pipeline that include: Acquisition, Transformation, Storage, Processing, Analysis and Visualization. Then, they have developed a framework that clarifies some options for Big Data in a cloud computing context in order to improve the strategic alignment of Big Data applications within the enterprise.

In [60], researchers have introduced a BDVC consisting of five main phases: data acquisition, data analysis, data curation, data storage and data usage. Moreover, they have identified some key stakeholders and challenges in the Big Data ecosystems in Europe. These challenges have been identified using their BDVC and they also include the impact of Big Data technology in the determination of data value, the reduction of languages fragmentation, intellectual property rights, laws, and policy practices between EU countries.

Hu et al. [34], have presented BDVC as a systematic framework, following a system-engineering approach, and decomposed into four sequential modules: Data generation, Data acquisition, Data storage, and Data analytics. According to the authors, data visualization is considered as a method of assistance for the analysis phase. Moreover, they have provided a literature survey on approaches and mechanisms used to support their BDVC.

Authors in [66] have presented a BDVC that follows the hierarchical pyramid Data-Information-Knowledge-Wisdom of R. Ackloff [67]. This chain is divided into three phases, called Data Sources, Preprocessing and Storing, and Processing and Visualization. It describes how to improve the value created as the chain phases progress to reach wisdom. Researchers in [68] has discussed Big Data Analytics methodologies that work in pairs with BDVCs. In this regard, authors have proposed a value chain, composed of multiple phases (Acquisition/Recording---Data Integration, Aggregation and Representation---Query Processing, Data Modeling and Analysis---Interpretation), performed to fulfill defined Big Data objectives.

Regarding education field, the authors in [69] have presented a study based on the concept of BDVC to identify schools with students that have learning disabilities and the risk to leave the education system prematurely, thereby transform information into insights and knowledge.

BDVC has also attracted much attention in smart grid area, Daki et al. [70] has presented Big Data architecture by providing several steps from data sources to visualization. The proposed architecture enables to manage different flux coming from the electricity grid and various producers, users or communicating devices or deliver added value in the consumption and distribution of energy systems to monitor its sustainability. In the same context, the article in [71], applied to energy visualization scenarios, discusses a new approach based on BDVC for the segmentation of big data systems in smart grids environment. On the other hand, to ensure control, efficiency and sustainability of the smart grid, the authors in [72] have required appropriate computational platforms and analytical capabilities to extract the value from sensitive huge data streams of smart grid meters.

Other researchers have introduced different concepts in the BDVC such as the monetization and data quality. In their study on Big Data Quality metrics, the authors in [73]

have relied on a BDVC, suitable for sentiment analysis. It consists of three main phases; ingestion, analysis and visualization of data. This BDVC aims to retrieve and process Big Social Data to provide valuable and reliable insights. Serhani et al. [74] have proposed a hybrid approach to manage Big Data quality across BDVC. The aim of this approach is to assess the quality of Big Data itself, then, to assess the quality during BDVC execution. In this contribution, the BDVC includes: Data collection, Pre-Big Data quality evaluation, Pre-processing quality evaluation, Post-Big Data quality evaluation, Processing and analytics quality evaluation and Visualization. To improve metadata management for Big Data systems, the authors in [75] have proposed data quality architecture, based on Immonen's work [76]. They have relied on BDVC, composed on several steps of data acquisition, loading and preprocessing, processing, storage, analysis, loading and transformation, and interfacing and visualization.

Regarding data monetization, authors in [61] have added a monetization phase to the commonly known BDVC lifecycle. The proposed BDVC consists of five phases: Creation, Storage, Processing, Consumption and Monetization. The monetization phase has been described as both the final step and the key input/output component of financial planning. To achieve monetization, the authors in [77] have suggested a theory-driven Big Data Analysis guidance process. They have introduced a light theory based on guidance through a succession of steps: acquisition, pre-processing, analytics and interpretation in order to safeguard the analytics process from pitfalls.

### **Analysis of findings**

It is important to mention that even if most of research contributions have considered BDVC models for different areas, there is no contribution that adopts an end-to-end model that carefully consider every step along the BDVC. In this subsection, we summarize the steps considered in each of the above BDVC models in a comparison way (Table 2). This comparison aims to analyze and to project differences between the adopted models.

Figure 3 shows the rate of use of each phase in the above-mentioned models, according to their importance and use cases. From this graph, we notice that the monetization of data, as an integrated phase in VCs, is rarely mentioned.

In light of previous models, we have designed, in the above section, a global, comprehensive and exhaustive model capable to regroup mentioned phases in Table 2. In this way, we give to researchers more latitude and space to detail and implement our model. This model, contrary to those mentioned before, consider every aspect along big data value chain allowing users to control and generate value of data in its different forms.

It is important to mention that Big Data-driven organizations, that use BDVC approaches, generate higher benefits than traditional organizations since, once data is generated, it can be mined multiple times and for different needs. This mined data can be exploited by others in its different forms or built on and exploited many times over. Thus, an increasing number of those organizations are claiming to monetize their data as exchangeable or salable services. For this, enterprises can implement an effective strategy for data to be monetized in the raw or service form. In the next Section, we present some of these data monetization strategies and related business models.

**Table 2 Comparison of different BDVC models**

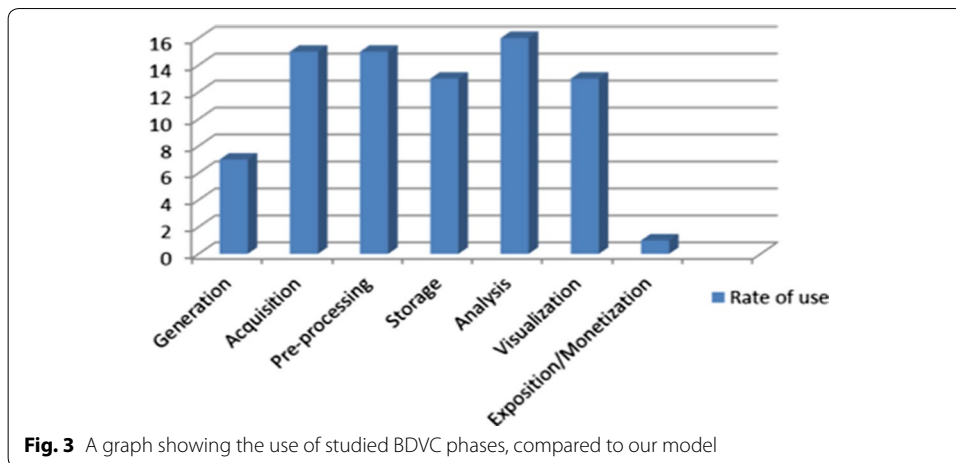
References	BDVC model steps & techniques	Context	Observations
[65]	Data discovery: collect/annotate-prepare-organize Data integration: integrate Data exploitation: analyze-visualize-make decisions	Creating value from disparate data and informing the enterprise decision-making	Storage is performed at the end of step "Data discovery"
[40]	Understanding of the application domain Selecting data and building input dataset Preprocessing and cleansing Transforming data Data mining Evaluating and interpreting patterns Visualization and feedback	Developing a Big Data processes understanding would enable business managers to improve business operations and ensure sustainable competitiveness	Analytical techniques were limited to data mining. The model targets business managers
[5]	Data acquisition/extraction Data transformation Data storage Data processing Data analysis Data visualization	Study of the possibilities of Big Data pipeline deployment in public, private or hybrid cloud	Data processing is performed after data storage
[60]	Data acquisition Data analysis Data curation Data storage Data usage	Generate value, model the high-level activities of an information system and be able to integrate an ecosystem	Data analysis is performed before the data pre-processing and storage
[34]	Data generation Data acquisition Data storage Data analysis	BDVC is presented as a framework that shows the evolution of data processing following a life cycle	Data visualization is considered as a support method for data analysis
[74]	Data collection Pre Big Data quality evaluation Pre-processing quality evaluation Post Big Data quality evaluation Processing and analytics quality evaluation. Visualization	Achieve a complete quality assessment of Big Data value chain through a hybrid model	The quality evaluation is applied in parallel with BDVC, in particular in pre-processing, processing and analysis phases
[66]	Data sources, types and accessibility Preprocessing and storing Processing and visualization	It is based on BDVC to achieve knowledge and wisdom	Phases too nested and strong coupling
[68]	Acquisition-recording Extraction-cleaning-annotation Integration-aggregation-representation Analysis-modeling Interpretation	BDVC model to realize goals in Big Data system	The datasets storage after pre-processing is not emphasized

**Table 2 (continued)**

References	BDVC model steps & techniques	Context	Observations
[75]	Data extraction Data loading and preprocessing Data processing Data storage Data analysis Data loading and transformation Data interfacing and visualization	Rely on BDVC to improve metadata management for Big Data systems	Repeated data loading presents a latency for real or near real time processing
[77]	Acquisition Pre-processing Analytics Interpretation	Theory-driven Big Data Analysis guidance process	Storage levels is not clear to assert the construction of the datasets for monetization
[61]	Creation: data capture Storage: warehousing Processing: data mining, fusion, and analytics Consumption: sharing Monetization	Rely on BDVC to realize data monetization, which is part of a global strategy, including different stakeholders	Monetization is the last phase of the chain, added after consumption
[71]	Data generation Data acquisition Data storing Data processing Data querying Data analytics Monitoring	Extracting value from Big Data systems in smart grids environment	The preprocessing step is not detailed in spite of its importance in a Big Data lifecycle
[69]	Data acquisition Data aggregation Data processing Data delivery	Discover ideas and knowledge about the governance sector	The approach adopted is based more on Big Data Analytics than BDVC
[70]	Data sources Data integration Data storage Data analytics Data visualization	Big Data architecture of smart grid to enable monitoring energy systems	Smart grid data are often unstructured. Preprocessing phase must be specified
[73]	Data ingestion Data analysis Data visualization	Application of Big Data metrics based on a BDVC adapted to sentiment analysis	This model cannot be generalized on different Big Data environments

### Towards Big Data monetization

The monetization concept aims to generate measurable economic benefits from both raw data and mined resources. This can generate value in mainly two ways : directly by selling or sharing data called explicit monetization, or indirectly by enhancing own data-based products known as implicit monetization [78]. It is thus very important for



organizations to adopt strategies and models to monetize data in an efficient way at the right moment. This involves the consideration of various aspects and pathways of data management that we will discuss in the next Subsections.

### The evolution of Big Data monetization

The monetization plays a key role in Big Data projects and has become essential to bring additional value [79–81]. Therefore, it is important to highlight the evolution of data monetization from the Business Intelligence (BI) era to the rise of Big Data, bringing an evolution in analytics capabilities [61, 82].

In the last decade, enterprises have become more oriented towards data visualization approaches. In that period, BI has become a strategic topic and was adopted by business and technology leaders as instrumental in driving industries. BI is the process of technology-driven data analytics tools that extracts usable and sharable information. These tools can then be used to create reports, dashboards, and data visualizations in order to make analysis results available to decision-makers [83]. During the 2000s, the monetization of these data was generally implicit, performed by a primarily descriptive analytics and reporting, which allows the extraction, processing and aggregation of production data for both internal use or internal exchange between the departments [84].

From 2010, with the rise of technologies like Big Data and IoT, the concept of data monetization has gained popularity. There are more and more data, especially with sensors and other web applications. This new mode has led many organizations to change their way of dealing with the data and started to integrate data from multiple external sources with internal data within data management platforms. A data management platform ensures an end-to-end integrity of all processes within the organization to provide new possibilities to access and exploit valuable information, and to enhance organization services and product [85].

To face Big Data challenges as explained above in Section 'Big Data characteristics', data management platforms include advanced analytics and computational capabilities, and also leverage cloud possibilities [61, 86] (Data storage and analysis in hybrid cloud). This kind of data management platforms enables building large information databases,



drive business decisions and provide data-driven products and services [87, 88]. Furthermore, it allows an explicit monetization mode to be introduced by selling data in different maturity levels [82]. That is, Big Data and data management platforms are becoming key controllers of a fast and agile delivery mode for creating and monetizing knowledge [64, 81].

#### **Pathways for Big Data Monetization business models**

In order to monetize data, many business models and pathways have been proposed. Next, we present some literature review in this regard.

Recently, Big Data Monetization has received attention not only from organizations but also from scientific research. In fact, several researches have proposed different business models, strategies and pathways to follow in order to monetize data in different contexts.

Researchers in [89], have highlighted six new ways, that differ from conventional sources of revenue, to monetize social networks, including side payments, affiliate programs, access controls, aggregation rules and integrated mobile platforms. Furthermore, authors have presented two other ways to create non-monetary value through social networks. First, Customer Interaction increases brand loyalty and improve customer feedback. Second, Knowledge Management facilitates interaction between employees.

In [82], Najjar and Kettinger have provided three pathways that can help companies considering data monetization initiatives. The plotted pathways are: (i) building high analytical capability based on mathematical and business analytical knowledge, (ii) building high technical data infrastructure by considering the hardware, software, and network aspects. Authors have also highlighted the importance of sharing data with suppliers in order to improve their analytical capabilities and also to avoid some analytical over costs.

In [90], authors have proposed a model based on several dimensions, related to the retail sector, to apply predictive and statistical analysis in a Big Data context. This allows predictive models to be designed and monetize the data value by understanding and mastering the elasticity of prices and promotions in retailing.

Researchers have exposed in [91] Data Monetization Journey of BBVA global financial group in various domains. It concerns the direct generation of benefits by selling information solutions to external customers, the improvement of core business operations that can provide large benefic value, the application of analytical processing to guide products and the economic impact benefit of data monetization-based projects. For this, new approaches have been pursued basing on the construction and consolidation of capabilities in data science, the acquisition of IT infrastructures, analytical capabilities, talents and skilled jobs.

In [92], authors have given an unified analytic framework for Big Data Monetization in telecom sector. This framework exploits Big Data analytic capabilities and takes advantage of the subscriber databases of large telecoms networks in order to study users' profiles and behaviors and also to look for like-minded communities. The proposed

contribution aims to monetize data by exploiting a word-of-mouth or viral marketing strategy.

Opher et al. [64] have presented a report that explains the emerging data economy with data monetization. Authors specify that IoT have made huge data assets in motion and offer systems able to transform business and improve the human experience. Thus, customer expectations and technological advancements will transform the supply chain to complex meshes ecosystems, based on suppliers, retailers, manufacturers and distributors interactions. Consumers and partners facilitate the creation of interconnected supply chains that allow high and direct collaboration, upstream or downstream, which requires orchestration rather than control procedures.

In healthcare sector, Visconti et al. [61], have considered monetization as the last phase of BDVC: Creation, Storage, Processing, Consumption. This last phase aims to simply transform added value into cash. Authors have also shown that processing data collected from heterogeneous sources, would allow via network platforms, to define new data sets, find correlations and hidden interpretations to ensure a maximum monetization profits.

In a creation Strategic Business Value context, the authors in [93] have provided a framework for value-based creation on Big Data Analytics. The researchers have found that firms are far from an efficient application data analytic and are not aligned for value creation. They have specified that enterprises must evaluate the strategic role of Big Data analysis and invest in quality data despite the challenges posed by the Big Data Vs complexity. Besides, the mediation between capacities and their realization of Big Data system remains necessary to improve strategic value targets.

To sum up, creating business models for monetizing Big Data can be articulated around four main axes: (i) Data extracted from customers' activities which could be in its raw format. (ii) Data providers that collect and sale primary and secondary data. (iii) Data aggregators that provide customers with aggregated services. (iv) Technical platforms, based on infrastructure, analysis, computing and cloud capabilities that enable to process, consume and share data.

Despite the importance of Big Data Monetization, few researches have considered the monetization concerns in a Big Data context. However, the monetization, both direct and indirect, has always been a guarantee for the performance and the organizations' durability. In fact, organizations are always looking for a return on investment on the deployment of BDVC. Therefore, owning the data and being able to monetize it: this is certainly the key success factor of tomorrow's leader, one that will allow everyone to innovate and create new business models. Thus, many efforts should be addressed to identify suitable strategies for an efficient data monetization approach.

## Conclusion

VC consists of analyzing information in order to organize processes that contribute directly or indirectly to creating added value for the organization. It is worth noting that, with the digitization of daily activities as well as how much the Internet is becoming popular, the way to process information has completely changed. In fact, organizations are now gathering and processing data from different sources in order to extract useful insights and hidden patterns. However, new data characteristics have emerged such

as Volume, Velocity, Variety, Veracity, Variability, Visualization and Value, requiring not only new technologies and infrastructures to process it, but also an efficient manner to manage and create value from those records. For this, new VC models have been established to help us process and cope with information challenges. Once processed, this data can offer multiple opportunities and generate maximum benefits for businesses; helping to better understand market trends, to make timely business decisions and also to produce valuable datasets to monetize by selling or sharing them. This data monetization will enable to unlock the value and reach completely data-driven processes.

In this paper, we review the background and the state-of-the-art of value chain evolution towards BDVC. First, we presented the general background and the history of both VC and DVC. Then, we presented Big Data characteristics that pushed organizations to adopt BDVC models. Moreover, we go through BDVC phases such as: Data Generation, Data Acquisition, Data Preprocessing, Data Storage, Data Analysis, Data Visualization and Data Exposition, by highlighting other proposed models that aims to define BDVC in different ways. Furthermore, we explore the related works that propose possible adaptations and enhancements of this kind of chains in different contexts. Finally, we discuss the evolution and the importance of Big Data Monetization by projecting various aspects and pathways to implement such vision.

During the setting up of this state-of-the-art, we have noticed that, to the best of our knowledge, there is not enough academic research in Big Data Monetization. Furthermore, the existing ones do not consider all big data aspects such as pre-processing, cleaning and other big data specificities. These specificities are key elements to monetize reliable value. For this purpose, our future work aims to consider an appropriate approach of BDVC in order to build an efficient and global model adapted to Big Data Monetization on a large scale.

#### **Abbreviations**

VC: Value Chain; DVC: Data Value Chain; BDVC: Big Data Value Chain; IoT: Internet of Things; KM: Knowledge Management; KVC: Knowledge Value Chain; URI: Uniform Resource Identifier; RDF: resource description framework; HTTP: hypertext transfer protocol; LDVC: Linked Data Value Chain; NoSQL: Not Only SQL; ERP: Enterprise Resource Planning; BI: Business Intelligence.

#### **Acknowledgements**

Not applicable.

#### **Authors' contributions**

AF has conducted a literature review and examined different models related to Big Data Value Chains as well as data monetization. AF has also proposed the global Big Data Value Chain and wrote the initial draft paper. IE has structured and adapted the manuscript. YG has provided guidelines and technical details, required to fulfill the study. AF, IE and YG have written, reviewed the manuscript and validate the study proposal and design. AA has critically reviewed and helped to prepare the final manuscript. All authors read and approved the final manuscript.

#### **Funding**

Not applicable.

#### **Availability of data and materials**

Not applicable.

#### **Competing interests**

The authors declare that they have no competing interests.

#### **Author details**

<sup>1</sup> Laboratoire Génie des Systèmes, Ecole Nationale des Sciences Appliquées de Kenitra, University of Ibn Tofail, Kenitra, Morocco. <sup>2</sup> Laboratoire des Systèmes de Télécommunications et Ingénierie de la Décision, University of Ibn Tofail, Kenitra, Morocco.

Received: 25 October 2019 Accepted: 23 December 2019

Published online: 08 January 2020

## References

- Porter ME. Clusters and the new economics of competition. *Harv Bus Rev.* 1998;76(6):77–90.
- Micek G. Competition, competitive advantage and clusters: the ideas of Michael Porter—edited by Robert Huggins & Hiro Izushi: book reviews. *Tijdschrift voor economische en sociale geografie.* 2012;103(2):250–2. <https://doi.org/10.1111/j.1467-9663.2012.00704.x>.
- Prajogo DI, McDermott P, Goh M. Impact of value chain activities on quality and innovation. *Int J Oper Prod Manage.* 2008;28(7):615–35. <https://doi.org/10.1108/01443570810881785>.
- Demchenko Y, de Laat C, Membrey P. Defining architecture components of the Big Data ecosystem. In: 2014 international conference on collaboration technologies and systems (CTS). IEEE, Minneapolis, MN, USA; 2014, p. 104–12. <https://doi.org/10.1109/CTS.2014.6867550>.
- Schmidt R, Mohring M. Strategic alignment of cloud-based architectures for Big Data. In: 2013 17th IEEE international enterprise distributed object computing conference workshops. IEEE, Vancouver, BC, Canada; 2013, p. 136–43. <https://doi.org/10.1109/EDOCW.2013.22>.
- McAfee A, Brynjolfsson E. Big data: the management revolution. *Harv Bus Rev.* 2012;90:9.
- Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br J Manage.* 2003;14(3):207–22. <https://doi.org/10.1111/1467-8551.00375>.
- Timmers P. Business models for electronic markets. *Electron Mark.* 1998;8(2):3–8. <https://doi.org/10.1080/10196789800000016>.
- Competitive advantage : creating and sustaining superior performance : with a new introduction. Free Press, Originally Published
- Holsapple CW, Singh M. The knowledge chain model: activities for competitiveness. *Expert Syst Appl.* 2001;20(1):77–98. [https://doi.org/10.1016/S0957-4174\(00\)00050-6](https://doi.org/10.1016/S0957-4174(00)00050-6).
- Carlucci D, Schiuma G. Knowledge asset value spiral: linking knowledge assets to company's performance. *Knowl Process Manage.* 2006;13(1):35–46. <https://doi.org/10.1002/kpm.243>.
- Chyi Lee C, Yang J. Knowledge value chain. *J Manage Dev.* 2000;19(9):783–94. <https://doi.org/10.1108/0262171000378228>.
- Kodó K, Hahn I. Literature review of the value grid model, 2017. <http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-33421>. Accessed 6 Dec 2019.
- Pil FK, Holweg M. Evolving from value chain to value grid. *MIT Sloan Manage Rev.* 2006;47(4):72–80.
- Universal Resource Identifiers: Contents. [https://www.w3.org/Addressing/URL/URI\\_Overview.html](https://www.w3.org/Addressing/URL/URI_Overview.html). Accessed 06 Dec 2019.
- RDF Schema 1.1. <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>. Accessed 06 Dec 2019.
- Data—W3C. <https://www.w3.org/standards/semanticweb/data/>. Accessed 06 Dec 2019.
- Latif A, Höfler P, Stocker A, Saeed AU, Wagner C. The linked data value chain: a lightweight model for business engineers. *Proc I-SEMANTICS.* 2009;2009:568–75.
- Peppard J, Rylander A. From value chain to value network. *Eur Manage J.* 2006;24(2–3):128–41. <https://doi.org/10.1016/j.emj.2006.03.003>.
- Stanoevska-Slabeva K, Talamanca C F, Thanos G A, Zsigri C. Development of a generic value chain for the grid industry. In: *Grid economics and business models*, vol. 4685, pp. 44–57. Springer, Berlin, Heidelberg; 2007. [https://doi.org/10.1007/978-3-540-74430-6\\_4](https://doi.org/10.1007/978-3-540-74430-6_4).
- Attard J, Orlandi F, Auer S. Data value networks: enabling a new data ecosystem. In: 2016 IEEE/WIC/ACM international conference on web intelligence (WI). IEEE, Omaha, NE, USA; 2016, p. 453–6. <https://doi.org/10.1109/WI.2016.0073>.
- Khan M A-u-d, Uddin M F, Gupta N. Seven V's of Big Data understanding Big Data to extract value. In: *Proceedings of the 2014 zone 1 conference of the American society for engineering education*. IEEE, Bridgeport, CT, USA; 2014, p. 1–5. <https://doi.org/10.1109/ASEEZone1.2014.6820689>.
- Anshari M, Lim SA. E-Government with Big Data enabled through smartphone for public services: possibilities and challenges. *Int J Public Admin.* 2017;40(13):1143–58. <https://doi.org/10.1080/01900692.2016.1242619>.
- Al-Sai Z A, Abualligah L M. Big data and E-government: a review. In: 2017 8th international conference on information technology (ICIT), p. 580–7. IEEE, Amman, Jordan; 2017. <https://doi.org/10.1109/ICITECH.2017.8080062>.
- Zainal N Z, Hussin H, Nazri M N M. Big data initiatives by governments—issues and challenges: a review. In: 2016 6th international conference on information and communication technology for the Muslim World (ICT4M), p. 304–9. IEEE, Jakarta, Indonesia; 2016. <https://doi.org/10.1109/ICT4M.2016.068>.
- Kshetri N. The emerging role of Big Data in key development issues: opportunities, challenges, and concerns. *Big Data Soc.* 2014;(2):205395171456422. <https://doi.org/10.1177/2053951714564227>.
- Barham H. Achieving competitive advantage through Big Data: a literature review. In: 2017 Portland international conference on management of engineering and technology (PICMET), p. 1–7. IEEE, Portland, OR 2017. <https://doi.org/10.23919/PICMET.2017.8125459>.
- Rabi T, Gómez-Villamor S, Sadoghi M, Muntés-Mulero V, Jacobsen H-A, Mankovskii S. Solving Big Data challenges for enterprise application performance management. *Proc VLDB Endow.* 2012;5(12):1724–35. <https://doi.org/10.14778/2367502.2367512>.
- Al Nuaimi E, Al Neyadi H, Mohamed N, Al-Jaroodi J. Applications of Big Data to smart cities. *J Internet Serv Appl.* 2015;6(1):25. <https://doi.org/10.1186/s13174-015-0041-5>.
- Agrahari A, Rao D. A review paper on Big Data: technologies, tools and trends. *Int Res J Eng Technol.* 2017;4(10):10.
- Victor N, Lopez D, Abawajy JH. Privacy models for Big Data: a survey. *Int J Big Data Intell.* 2016;3(1):61. <https://doi.org/10.1504/IJBDI.2016.073904>.

32. Saggi MK, Jain S. A survey towards an integration of Big Data analytics to big insights for value-creation. *Inform Process Manage*. 2018;54(5):758–90. <https://doi.org/10.1016/j.ipm.2018.01.010>.
33. El Alaoui I, Gahi Y, Messoussi R. Full consideration of Big Data characteristics in sentiment analysis context. In: 2019 IEEE 4th international conference on cloud computing and Big Data analysis (ICCCBDA). IEEE, Chengdu, China; 2019, p. 126–30. <https://doi.org/10.1109/ICCCBDA.2019.8725728>.
34. Han Hu, Wen Yonggang, Chua Tat-Seng, Li Xuelong. Toward scalable systems for Big Data analytics: a technology tutorial. *IEEE Access*. 2014;2:652–87. <https://doi.org/10.1109/ACCESS.2014.2332453>.
35. Adnan K, Akbar R. An analytical study of information extraction from unstructured and multidimensional Big Data. *J Big Data*. 2019;6(1):91. <https://doi.org/10.1186/s40537-019-0254-8>.
36. Erl T, Khattak W, Buhler P. Big data fundamentals: concepts, drivers & techniques. The Prentice Hall service technology series from Thomas Erl. Prentice Hall ; Service Tech Press, Boston : [Vancouver, BC] 2016. OCLC: ocn918932107.
37. Rehman M H u, Chang V, Batool A, Wah TY. Big data reduction framework for value creation in sustainable enterprises. *Int J Inform Manage*. 2016;36(6):917–28. <https://doi.org/10.1016/j.ijinfomgt.2016.05.013>.
38. Chen M, Mao S, Liu Y. Big data: a survey. *Mob Netw Appl*. 2014;19(2):171–209. <https://doi.org/10.1007/s11036-013-0489-0>.
39. Siddiqi A, Hashem IAT, Yaqoob I, Marjani M, Shamsirband S, Gani A, Nasaruddin F. A survey of Big Data management: taxonomy and state-of-the-art. *J Netw Comput Appl*. 2016;71:151–66. <https://doi.org/10.1016/j.jnca.2016.04.008>.
40. Rajpurohit A. Big data for business managers—bridging the gap between potential and value. In: 2013 IEEE international conference on Big Data. IEEE, Silicon Valley, CA, USA; 2013, p. 29–31. <https://doi.org/10.1109/BigData.2013.6691794>.
41. ur Rehman MH, Liew CS, Abbas A, Jayaraman PP, Wah TY, Khan SU. Big data reduction methods: a survey. *Data Sci Eng*. 2016;1(4):265–84. <https://doi.org/10.1007/s41019-016-0022-0>.
42. Santhanam T, Padmavathi M S. Comparison of K-Means clustering and statistical outliers in reducing medical datasets. In: 2014 international conference on science engineering and management research (ICSEMR). IEEE, Chennai, India; 2014, p. 1–6. <https://doi.org/10.1109/ICSEMR.2014.7043602>.
43. Salmon J, Harmany Z, Deledalle C-A, Willett R. Poisson noise reduction with non-local PCA. *J Math Imag Vision*. 2014;48(2):279–94. <https://doi.org/10.1007/s10851-013-0435-6>.
44. Zhai Y, Ong Y-S, Tsang IW. The emerging "big dimensionality". *IEEE Comput Intell Mag*. 2014;9(3):14–26. <https://doi.org/10.1109/MCI.2014.2326099>.
45. García S, Ramírez-Gallego S, Luengo J, Benítez JM, Herrera F. Big data preprocessing: methods and prospects. *Big Data Anal*. 2016;1(1):9. <https://doi.org/10.1186/s41044-016-0014-0>.
46. Ahamed B B, Ramkumar T, Hariharan S. Data integration progression in large data source using mapping affinity. In: 2014 7th international conference on advanced software engineering and its applications. IEEE, Hainan Island, China; 2014, p. 16–21. <https://doi.org/10.1109/ASEA.2014.11>.
47. Kasim H, Hung T, Li X. Data value chain as a service framework: for enabling data handling, data security and data analysis in the cloud. In: 2012 IEEE 18th international conference on parallel and distributed systems. IEEE, Singapore, Singapore; 2012, p. 804–9. <https://doi.org/10.1109/ICPADS.2012.131>
48. Wu D, Sakr S, Zhu L. Big data storage and data models. In: Zomaya AY, Sakr S, editors. *Handbook of Big Data technologies*. Cham: Springer; 2017. p. 3–29. [https://doi.org/10.1007/978-3-319-49340-4\\_1](https://doi.org/10.1007/978-3-319-49340-4_1).
49. Wu D, Sakr S, Zhu L. Big data programming models. In: Zomaya AY, Sakr S, editors. *Handbook of Big Data technologies*. Cham: Springer; 2017. p. 31–63. [https://doi.org/10.1007/978-3-319-49340-4\\_2](https://doi.org/10.1007/978-3-319-49340-4_2).
50. Dutta K. Distributed computing technologies in Big Data analytics. In: Mazumder S, Singh Bhadoria R, Deka GC, editors. *Distributed computing in Big Data analytics*. Cham: Springer; 2017. p. 57–82. [https://doi.org/10.1007/978-3-319-59834-5\\_4](https://doi.org/10.1007/978-3-319-59834-5_4).
51. Storey VC, Song I-Y. Big data technologies and management: what conceptual modeling can do. *Data Knowl Eng*. 2017;108:50–67. <https://doi.org/10.1016/j.datak.2017.01.001>.
52. Cao J, Chawla S, Wang Y, Wu H. Programming platforms for Big Data analysis. In: Zomaya AY, Sakr S, editors. *Handbook of Big Data technologies*. Cham: Springer; 2017. p. 65–99. [https://doi.org/10.1007/978-3-319-49340-4\\_3](https://doi.org/10.1007/978-3-319-49340-4_3).
53. Shi-Nash A, Haroon DR. Data analytics and predictive analytics in the era of Big Data. In: Geng H, editor. *Internet of things and data analytics handbook*. Hoboken: Wiley; 2016. p. 329–45. <https://doi.org/10.1002/9781119173601.ch19>.
54. Stimmel CL. *Big data analytics strategies for the smart grid*. Boca Raton: CRC Press; 2015.
55. Heidari S, Alborzi M, Radfar R, Afsharkazemi MA, Rajabzadeh Ghatari A. Big data clustering with varied density based on MapReduce. *J Big Data*. 2019;6(1):77. <https://doi.org/10.1186/s40537-019-0236-x>.
56. Ait-Mlouk A, Agouti T, Gharnati F. Mining and prioritization of association rules for Big Data: multi-criteria decision analysis approach. *J Big Data*. 2017;4(1):42. <https://doi.org/10.1186/s40537-017-0105-4>.
57. Becker T. Big data usage. In: Cavanillas JM, Curry E, Wahlster W, editors. *New horizons for a data-driven economy*. Cham: Springer; 2016. p. 143–65. [https://doi.org/10.1007/978-3-319-21569-3\\_8](https://doi.org/10.1007/978-3-319-21569-3_8).
58. Alexandru A, Alexandru C, Coardos D, Tudora E. Big data: concepts, technologies and applications in the public sector. *Int J Comput Electr Autom Control Inform Eng*. 2016;10:1629–35. <https://doi.org/10.5281/ZENODO.1126762>.
59. Chen P. Visualization of real-time monitoring datagraphic of urban environmental quality. *EURASIP J Image Video Process*. 2019;2019(1):42. <https://doi.org/10.1186/s13640-019-0443-6>.
60. Curry E. The Big Data value chain: definitions, concepts, and theoretical approaches. In: Cavanillas JM, Curry E, Wahlster W, editors. *New horizons for a data-driven economy*. Cham: Springer; 2016. p. 29–37. [https://doi.org/10.1007/978-3-319-21569-3\\_3](https://doi.org/10.1007/978-3-319-21569-3_3).
61. Moro Visconti R, Larocca A, Marconi M. Big data-driven value chains and digital platforms: from value co-creation to monetization. *SSRN Electron J*. 2017; <https://doi.org/10.2139/ssrn.2903799>.
62. Chen Chiang. Storey: business intelligence and analytics: from Big Data to big impact. *MIS Q*. 2012;36(4):1165. <https://doi.org/10.2307/41703503>.

63. Dong Xinhua, Lu Ruixuan, He Heng, Zhou Wanwan, Xue Zhengyuan, Hao Wu. Secure sensitive data sharing on a Big Data platform. *Tsinghua Sci Technol.* 2015;20(1):72–80. <https://doi.org/10.1109/TST.2015.7040516>.
64. Opher A, Chou A, Onda A. The rise of the data economy: driving value through internet of things data monetization. Somers: IBM Corporation; 2016. p. 16.
65. Miller HG, Mork P. From data to decisions: a value chain for Big Data. *IT Prof.* 2013;15(1):57–9. <https://doi.org/10.1109/MITP.2013.11>.
66. Jony RI, Rony RI, Rahman M, Rahat A. Big data characteristics, Value chain and challenges. In: Proceedings of the 1st international conference on advanced information and communication technology. Bangladesh; 2016.
67. Ackoff RL. From data to wisdom. *J Appl Syst Anal.* 1989;16(1):3–9.
68. Ramannavar M, Sidnal N S. Big data and analytics—a journey through basic concepts to research issues. In: Suresh LP, Panigrahi BK (eds.) Proceedings of the international conference on soft computing systems, vol. 398, p. 291–306. Springer, New Delhi; 2016. [https://doi.org/10.1007/978-81-322-2674-1\\_29](https://doi.org/10.1007/978-81-322-2674-1_29).
69. Petrova-Antonova D, Georgieva O, Ilieva S. Modelling of educational data following Big Data value chain. In: Proceedings of the 18th international conference on computer systems and technologies—CompSysTech'17. ACM Press, Ruse, Bulgaria; 2017, p. 88–95. <https://doi.org/10.1145/3134302.3134335>.
70. Daki H, El Hannani A, Aqqal A, Haidine A, Dahbi A. Big data management in smart grid: concepts, requirements and implementation. *J Big Data.* 2017;4(1):13. <https://doi.org/10.1186/s40537-017-0070-y>.
71. Munshi AA, Mohamed YA-RI. Big data framework for analytics in smart grids. *Electr Power Syst Res.* 2017;151:369–80. <https://doi.org/10.1016/j.epsr.2017.06.006>.
72. Zhang Y, Huang T, Bompard EF. Big data analytics in smart grids: a review. *Energy Inform.* 2018;1(1):8. <https://doi.org/10.1186/s42162-018-0007-5>.
73. El Alaoui I, Gahi Y, Messoussi R. Big data quality metrics for sentiment analysis approaches. In: Proceedings of the 2019 international conference on Big Data engineering (BDE 2019)—BDE 2019. ACM Press, Hong Kong, Hong Kong; 2019, p. 36–43. <https://doi.org/10.1145/3341620.3341629>.
74. Serhani M A, El Kassabi H T, Taleb I, Nujum, A. An hybrid approach to quality evaluation across Big Data value chain. In: 2016 IEEE international congress on Big Data (BigData Congress). IEEE, San Francisco, CA; 2016, pp. 418–425. <https://doi.org/10.1109/BigDataCongress.2016.65>.
75. Pääkkönen P, Jokitulppo J. Quality management architecture for social media data. *J Big Data.* 2017;4(1):6. <https://doi.org/10.1186/s40537-017-0066-7>.
76. Immonen A, Paakkonen P, Ovaska E. Evaluating the quality of social media data in Big Data architecture. *IEEE Access.* 2015;3:2028–43. <https://doi.org/10.1109/ACCESS.2015.2490723>.
77. Elragal A, Klischewski R. Theory-driven or process-driven prediction? Epistemological challenges of Big Data analytics. *J Big Data.* 2017;4(1):19. <https://doi.org/10.1186/s40537-017-0079-2>.
78. Wixom BH, Ross JW. How to monetize your data. *MIT Sloan Manage Rev.* 2017;58(3):9–13.
79. Li WC, Nirei M, Yamana K. Value of data: there's no such thing as a free lunch in the digital economy. Technical report, Research Institute of Economy, Trade and Industry (RIETI); 2019.
80. Davenport TH, Bean R. Big companies are embracing analytics, but most still don't have a data-driven culture. *Harv Bus Rev.* 2018;6:4.
81. Bughin J. Reaping the benefits of Big Data in telecom. *J Big Data.* 2016;3(1):14. <https://doi.org/10.1186/s40537-016-0048-1>.
82. Najjar M, Kettinger W. Data monetization: lessons from a retailer's journey. *MIS Q Executive.* 2013;12(4):189–98.
83. Watson HJ, Wixom BH. The current state of business intelligence. *Computer.* 2007;40(9):96–9. <https://doi.org/10.1109/MC.2007.331>.
84. Turban E, editor. Business intelligence: a managerial approach. Upper Saddle River: Pearson Prentice Hall; 2008 OCLC: ocm85623158.
85. Walker R. From Big Data to big profits: success with data and analytics. New York: Oxford University Press; 2015 OCLC: ocn907178027.
86. Mazumdar S, Seybold D, Kritikos K, Verginadis Y. A survey on data storage and placement methodologies for Cloud-Big Data ecosystem. *J Big Data.* 2019;6(1):15. <https://doi.org/10.1186/s40537-019-0178-3>.
87. Liu C-H, Chen C-L. A review of data monetization: strategic use of Big Data. In: The fifteenth international conference on electronic business (ICEB 2015); 2015, p. 7.
88. Hunke F, Engel C. Utilizing data and analytics to advance service: towards enabling organizations to successfully ride the next wave of servitization. In: Satzger G, Patrício L, Zaki M, Kühl N, Hottum P, editors. Exploring service science, vol. 331. Cham: Springer; 2018. p. 219–31. [10.1007/978-3-030-00713-3\\_17](https://doi.org/10.1007/978-3-030-00713-3_17).
89. Gomez-Arias JT, Genin L. Beyond monetization: creating value through online social networks. *IJEBM.* 2009;7:79–85.
90. Bradlow ET, Gangwar M, Kopalle P, Voleti S. The role of Big Data and predictive analytics in retailing. *J Retail.* 2017;93(1):79–95. <https://doi.org/10.1016/j.jretai.2016.12.004>.
91. Alfaro E, Bressan M, Girardin F, Murillo J, Someh I, Wixom BH. BBVA's data monetization journey. *MIS Q Executive.* 2019;18(2):117–28. <https://doi.org/10.17705/2msqe.00011>.
92. Cao H, Dong W, Liu L, Ma C, Qian W, Shi J, Tian C, Wang Y, Konopnicki D, Shmueli-Scheuer M, et al. SoLoMo analytics for telco Big Data monetization. *IBM J Res Dev.* 2014;58(5/6):9–1913. <https://doi.org/10.1147/JRD.2014.2336177>.
93. Grover V, Chiang RHL, Liang T-P, Zhang D. Creating strategic business value from Big Data analytics: a research framework. *J Manage Inform Syst.* 2018;35(2):388–423. <https://doi.org/10.1080/07421222.2018.1451951>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.