Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 151 (2019) 614–621

Procedia Computer Science

www.elsevier.com/locate/procedia

The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40)
April 29 – May 2, 2019, Leuven, Belgium

Towards a Data Governance Framework for Third Generation Platforms

Juan Yebenes, Marta Zorrilla*

ISTR group, Universidad de Cantabria,Avda. Los Castros s/n, Santander, 39005, Spain

Abstract

The fourth industrial revolution considers data as a business asset and therefore this is placed as a central element of the software architecture (data as a service) that will support the horizontal and vertical digitalization of industrial processes. The large volume of data that the environment generates, its heterogeneity and complexity, as well as its reuse for later processes (e.g. analytics, IA) requires the adoption of policies, directives and standards for its right governance. Furthermore, the issues related to the use of resources in the cloud computing must be taken into account with the aim of meeting the requirements of performance and security of the different processes. This article, in the absence of frameworks adapted to this new architecture, proposes an initial schema for developing an effective data governance programme for third generation platforms, that means, a conceptual tool which guides organizations to define, design, develop and deploy services aligned with its vision and business goals in I4.0 era.

Keywords: Data governance; Industry 4.0; data bus architecture; cloud computing; IoT; big data

1. Introduction

The fourth industrial revolution or Industry 4.0 deals with the transformation of production processes based on the use of the abundant information available in each phase of the production, logistics and consumption chains of an industrial sector [1]. The fourth industrial revolution faces a change in productive model that is based on the ubiquit

* Corresponding author. Tel.: +34-942-202063; fax: +34-942-201411.
E-mail address: marta.zorrilla@unican.es

10.1016/j.procs.2019.04.082

and connectivity of data, people, processes, services and cyber-physical systems, as if it were a social network in which, all the actors (network nodes) exchange and exploit the information generated at each level of the architecture (cyber-physical, intermediation and application level), which serves to achieve the required operation as well as a qualitative improvement in the automation and optimization of the industrial processes that, even, can lead to the creation of collaboration and industrial innovation ecosystems (horizontal and vertical integration).

To achieve the objectives that Industry 4.0 requires, it is necessary to develop and adopt methods, technologies and tools aimed at managing the following characteristics of industrial processes [2], such as the need to have high computing power in any environment (Ubiquitous & Mobile Computing), the capability to deal with huge volumes of data in real time and process them without the need to store it (Big Data), the dynamic scaling of the computing power according to workload changes (Cloud & Fog Computing), the reactive interaction of applications with intelligent environments and social networks (IE & IoT) and the use of computing paradigms based on artificial intelligence and machine learning (AI).

Although there are many architectural alternatives, according to [3], the so-called third generation platforms are suitable for answering to the technological challenge that Industry 4.0 requires. These are distributed and scalable architectures that are dynamically dimensioned based on the volume of computing that is required. Generally, these are comprised of a cluster of virtual or physical nodes, communication networks and middleware, which are managed under two paradigms:

- DaaS (Data as a Service): Data constitutes the highest level of the software architecture. The type of data or topic represents not only the business information that is assigned to it but also defines the non-functional requirements with which the former can be managed (persistence, durability, availability, security, integrity, etc.) and, that generally, are defined before the applications with the aim of enhancing the intercompany development of components. The processing tasks are decoupled software components that are executed in the platform nodes and interact between themselves by means of a publisher/subscriber paradigm of the topics defined by the applications following a data pipeline pattern [4][5]. These represent the internal control flow of the business transactions. Applications thus are designed as a workflow composed of a directed acyclic graph of processing tasks that are executed when certain patterns in the data environment occur.
- PaaS (Platform as a Service): Platforms are comprised by clusters of physical computational resources available in the working environment (fog) and /or external virtual computational resources (cloud) that can be dynamically hired according to the workload to be dealt with.

On the other hand, this paradigm shifts in the building of software systems, in which data is the backbone that supports and facilitates communication between the different connected elements (sensors, actuators, processes, services and people) presents new challenges for its management and good government. Data is an asset and must be necessarily governed to extract more value from them [6]. This means data management policies must be well defined and adhered to by validating data usage, quality, privacy and compliance (e.g. GDPR, HIPAA, SOX, etc.).

Although most organizations carry out some degree of data governance policies [7], Big Data and cloud computing add extra complexity as a result of inter-company collaboration, regulatory aspects and agreements of services hired to third parties [8]. This along with the specificity of the industrial environment, lead us to propose an effective and flexible Data Governance (DG hereinafter in this document) framework for next generation platforms after checking, by means of a systematic review, the lack of a conceptual tool about it.

The paper is organized as follows. Section 2 briefly describes the systematic review performed. Section 3 describes the proposed framework. Finally, Section 4 summarizes the final remarks and comments on our next future works.

2. Systematic review

This section gathers the results achieved as a consequence of the systematic literature review (SLR) performed in October 2018 with the aim of examining the most relevant published work about data governance for Industry 4.0 (hereinafter I4.0) and more precisely for third generation platforms (hereinafter 3GP). The SLR was carried out as described by Kitchenham and Charter [9].

The study attempted to address the following research questions: 1) What is the state-of-the-art of research in data/information governance in I4.0, in particular, and 3GP, in general? 2) What frameworks have been proposed for data/information governance in the I4.0, in particular, and the 3GP, in general?

In this study, we reviewed all published work between the year 2008 until 2018, written in English and in Spanish, in the following data bases: Scopus, Web of Science, Engineering Village, ScienceDirect, IeeeXplore, Abi Collection (Proquest), ACM, and TESEO. Our research included academic (journals, conferences, theses and reports) and practice-oriented sources from industry associations, software vendors and analysts.

The search is first performed on the publication title, abstract and keywords. The query carried out to answer our two question were:

Question 1: (TITLE-ABS-KEY("data governance" OR "information governance") AND TITLE-ABS-KEY(("industry 4.0" OR "smart factory" OR "smart manufacturing"))

Question 2: (TITLE-ABS-KEY("data governance" OR "information governance") AND TITLE-ABS-KEY("3rd platform" OR "third platform" OR "3rd generation platform" OR "third generation platform" OR "3rd generation platforms" OR "third generation platforms" OR "next-generation platform" OR "next-generation platforms"))

The result of the SLR shows that there is very little relevant literature on DG in the I4.0 environment or 3GP. In fact, we only found one document [10] that fits the questions asked and is suitable for our research. Therefore, the search was extended towards DG in the main technologies that support I4.0 and 3GP (IoT, cloud computing, data lakes, real-time data processing, cyber-physical systems and Big Data).

Table 1. Number of references found about data governance for I4.0, 3GP or related technologies.

Documents on DG / IG for	I4.0	3GP	Related Technologies	TOTAL
Documents found	5	1	824	830
Documents to review	0	1	65	66
Documents selected	0	1	21	22

As a result of this new search (see Table 1), a total of 830 documents were found which met the search criteria. After analyzing the tittle, abstract, and keywords, 66 documents were selected and reviewed in depth in order to ensure its validity for our scope. Finally, 40 references were discarded and 22 selected as suitable for our research goal (see Table 2). Ten out of 22 are focused on Cloud Computing, five are addressed to Data Lakes, four to Big Data, two about IoT and finally only one tackles third generation platforms.

From the analysis of these documents, we can conclude that, although there are specific models and proposals that address the problem of data governance in Cloud, Big Data, Data Lake, etc., we have not found a global or complete approach that encompasses the governance of data in the environment of I4.0. Nevertheless, we extracted some new challenges and risks which raise as a consequence of the adoption of these new technologies involved in I4.0 and that require the definition of specific polices and standards for their correct government. Thus, regarding IoT, we should consider security, scalability and interoperability issues due to an extremely large number of heterogeneous devices and a variety of protocols, domains and applications, also data quality issues (accuracy, confidence completeness, data volume and timeliness) as a consequence of the speed and infrastructure through which data are offered [31]. Other important aspect is the lack of a governance and security framework that can be effectively applied in the IoT-Cloud environment [22]. Also, cloud computing adds new issues regarding Deployment Models, Service Delivery Models, and Service Level Agreement (SLA) [26]. Moreover, the use of 3GP also implies processing large data sets, unstructured data, or low latency data which made more necessary a shift in data governance mind-sets in other to clarify the context, importance, and associations of data, and thus to ensure the data's reliability and security [10].

Table 2. Documents selected as related with the protocol search.

Tittle	Year	Source Area (*)	Domain
A conceptual framework for cloud data governance-driven decision making [11]	2017	A	Cloud Computing
A Conceptual Framework for Designing Data Governance for Cloud Computing [12]	2016	A	Cloud Computing
A Data Quality in Use model for Big Data [13]	2015	A	Cloud Computing
Accountability for Data Governance in Cloud Ecosystems [14]	2013	A	Cloud Computing
Accountability, Risk and Trust in Cloud Services Towards an Accountability-based Approach to Risk and Trust Governance [15]	2014	A	Cloud Computing
Aplicación de Cloud Data Lakes en la industria [16]	2016	P	Data Lakes
Bringing Accountability to the Cloud: Addressing Emerging Threats and Legal Perspectives [17]	2013	A	Cloud Computing
Case Study: Implementing Data Governance in Data Lakes and Big Data [18]	2017	P	Data Lakes
Cloud Data Governance Maturity Model [19]	2017	A	Cloud Computing
Data Governance: A conceptual framework in order to prevent your Data Lake from becoming a Data Swamp [20]	2015	A	Data Lakes
Data governance for Hadoop enabled enterprises [21]	2018	P	Big Data
Data governance for next-generation platforms [10]	2018	P	3GP
Data Governance for Security in IoT & Cloud Converged Environments [22]	2016	A	IoT
Data Governance Framework for Big Data Implementation with a Case of Korea [23]	2017	A	Big Data
Data lakes: purposes, practices, patterns and platforms [24]	2017	P	Data Lakes
Datameer Big Data Governance [25]	2015	P	Big Data
Development and Evaluation of a Holistic Framework and Maturity Assessment Tools for Data Governance in Cloud Computing Environments [26]	2018	A	Cloud Computing
Eliciting metrics for accountability of cloud systems [27]	2016	A	Cloud Computing
Governance of Big Data: perspective and issues [28]	2013	A	Big Data
Interoperability analysis of accountable data governance in the cloud [29]	2014	A	Cloud Computing
The Journey Continues. From Data Lake to Data-Driven Organization [30]	2018	P	Data Lakes
Toward data governance in the internet of things [31]	2018	A	IoT

(*) A = Academia, P = Practice-Oriented

3. Data Governance Framework for 3rd generation platforms (3GP)

It is important, before describing our proposed framework, to define the context of DG. Regrettably, there are a few definitions [32] [33] [34] [35] [6] but none of them can be considered official [26]. In fact, a standard definition of the term "Data Governance" can be found neither in the research community nor in the practitioners' community dealing with information systems [35].

We adopted the one proposed by [35] who defined DG as "A companywide framework for assigning decision-related rights and duties in order to be able to adequately handle data as a company asset", understanding assets, as data that has value or potential value for the company and so that should be documented. We refine this definition with the one proposed by DAMA [32] that defines DG as "the exercise of authority, control, and shared decision-

making (planning, monitoring and enforcement) over the management of data assets". In short, we can say that DG entails exercising authority, control and consensual and communicated decision-making over the management of data assets.

More specifically, DG must specify who has the decision-making rights and the responsibilities in the processes related to the information, as well as establishing a consensual model (framework) under which these processes are executed, describing who, what, when and under what circumstances and what methods should be used. In addition, it must orchestrate (and therefore affect) stakeholders, processes and technology.

DG also must be coherent with the mission, strategy, norms and culture of the organization so that it can manage its data as a strategic asset, providing quality control and protection in its access, management, supervision and maintenance, with the aim at getting more value from company data and turn it into a competitive advantage.

Traditional DG, as stated by [26], meets the requirements of companies and organizations with a traditional IT structure, which mainly involves structured data that are managed and stored on-premise but do not take into account the impact and new requirements derived from the use of a data-centered architecture and big data technologies. Therefore, this work aims to define a simple DG framework, maintainable and extendible, which allows to undertake DG with a global approach in an I4.0 environment supported on 3GP. Given that the architecture is data-oriented, we consider that the data life cycle is the central element of our proposal, supported on three pillars: an adequate metadata management, data quality and risk management (security and privacy) and all of it governed by policies directed towards company goals. Applications stop being developed as isolated processes, but they are built as the concatenation of tasks on data flows that can be easily parallelized in computing nodes and shared/used, in turn, by other processes. Therefore, the incorporation of these architectures also requires including activities aimed at the management of change and culture in the environment, sharing responsibilities and addressing principles and ethical aspects. Thus, our framework (see Fig. 1) identifies the domain areas to be governed, establishes the governance body as well as their decision rights and responsibilities on the domain areas of DG.

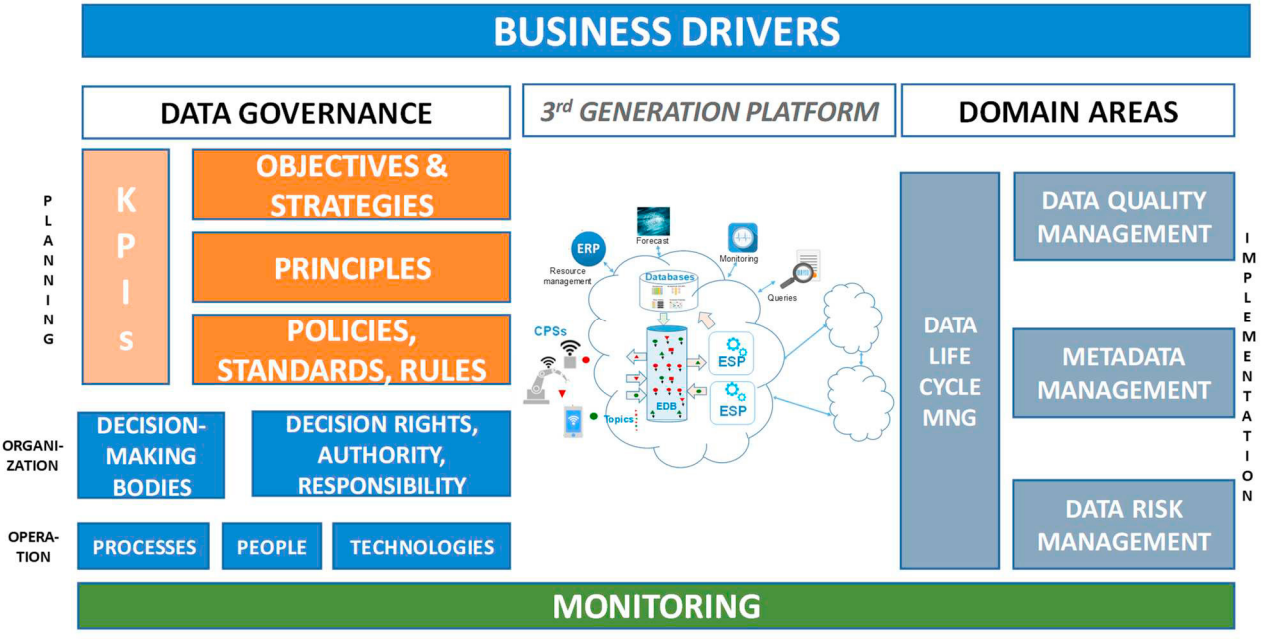

Fig. 1. DG framework for Industry 4.0

Planning.

The objectives and strategies of our DG program as well as the principles that explain and clarify how DG supports the structure, culture and goals of the company must be first specified, which must be coherent and aligned with the business objectives and strategies (business drivers). Next, data policies and standards that materialize the established principles and that govern the DG program must be established. DG policies must explicit the organization's strategic direction and high-level descriptions of both the desired behavior over the domain areas that is trying to achieve and

the desired outcome. So, policies must cover all data issues arising from the use of 3GP and I4.0, such as data security and risk, metadata management and data quality. Thus, policies must specify how data is governed at the different levels of vertical integration of I4.0 (edge, fog and cloud), as well as horizontal integration along the company's value chain. For instance, cloud architecture, implies that the defined policies must be implemented taking into account the third parties involved such as Cloud Provider, Cloud Carrier, Cloud Broker, etc. This leads to sign service level agreements (SLAs) between cloud actors [26], and consequently, general criteria and KPIs for internal control and establishing SLAs must be also defined at this stage.

These also imply defining and implementing policies that can be applied through different tools throughout the process of data management and that allow to verify how the data moves and is transformed through the system, meeting with the compliance and normative requirements (both internal and external), identifying the changes that have occurred in the data and who have carried them out in order to reduce the risk of leaks of sensitive information and to avoid the violation of privacy, thus promoting responsibility for the quality of the data [23].

Standards define the rules required to ensure that policies are properly implemented, so data fits its purpose. Processes and procedures provide details about how standards must be executed. Typical standards include specifications and directives about metadata management (ISO-11179, CWM, DCMI), data quality (ISO-8000, ISO-25012), data security (ISO-27001), among others. In short, it is about establishing how the data will be managed.

Organization.

Next, the required organization and structure to implement the DG program must be established. The aim is to identify the stakeholders involved in decision-making on data and its management at different levels of integration. Thus, it is necessary to define the roles and assign them rights on decision-making, responsibility and authority in the governance and management of the data.

To define the organization, we have taken, as a guide, part of the structure proposed by IBM [36], which distinguishes between the roles that govern and those that are governed. Among the former, the Executive Sponsor (ES), is the person or department that drives the need, funding, supervision and guidance of the DG initiative, clarifies and defines the scope of the DG initiative and helps to establish milestones and goals for completion, as well as to ensure compliance with all data laws and regulations. Next, the Data Governance Council (DGC) is comprised of a group of senior representatives from all business areas. Both, DGC and ES, set the objectives, as well as the general principles and policies of the DG program. Finally, the Data Governance Office (DGO) is the body responsible for providing and setting the data content, definition, context and associated definitions, rules, policies, and specifications for data content, data linage, data security and data quality. This is also responsible for ensuring that operational teams develop and implement the appropriate processes and procedures in the different data management areas. Experts in defining and hiring SLA must be included.

Regarding the roles that are governed, we consider the Business Data Steward, primarily accountable for the data owned by their business area. The Domain Data Steward is in charge of information assets and responsible for managing and maintaining its definition, implementation and documentation since a conceptual point of view. Lastly, the Technical Data Steward is responsible for the use, custody, transport, integration, quality and storage of information and its specifications. These roles are also responsible for supervising that the objectives of DG are met in those aspects of their competence that are developed outside the organization, as in the case of Cloud Computing. This requires establishing SLAs with cloud actors for those aspects that need to be monitored.

Operation.

One of the biggest challenges is transforming corporate data policies into processes and procedures that ensure its compliance. It is about defining how these policies and standards are implemented, translating them into processes and activities, that means, describing what the organization should do and how, with what technologies, models and languages suitable for a 3GP architecture. In this sense, the special characteristics of 3GP architecture makes the processes automation a fundamental element for DG in these environments, mainly regarding processes related to metadata.

It should be also considered that 3GP need people with profiles that are capable of evolving with the new technologies and dealing with the changes in the new way of operating with data and developing the automation of industrial processes. We must keep in mind that most data will be captured, stored and even analyzed in the lower levels of the network (edge computing layer), sometimes with the need for real-time processing, which will require specific tools for the management of quality, security and metadata, in real time, at different levels of integration.

Implementation.

Digital transformation in Industry supported in modern platforms involves processing large data sets mainly unstructured data and low latency data, which makes more difficult to understand the context, importance, and associations of data. Moreover, applications developed on 3GP require hiring of processing and storage services in the cloud to satisfy availability, latency and throughput requirements (among others) and all of this under demanding security and regulatory conditions. So, in terms of domain areas of DG, we will mainly focus on three issues: **data security and risks management**, which will require special attention when dealing with distributed environments and most data coming from of a large number of heterogeneous devices. Another important aspect is to guarantee the **data quality** (accuracy, completeness, consistency, credibility, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability, recoverability, timeliness, relevance, reliability, etc. [13]), that is affected by a greater quantity and complexity of the data (structured, semi-structured and unstructured), as well as the need to process low latency data. This leads to the need to manage a greater variety of **metadata** assets and develop more complex processes associated with them, in order to capture all the information we need for each type of data: data sources, means of creation, purpose, time and date of creation, applications using the data, data ingestion rate, sensitivity level, level of importance for the business, the useful life of the data, where it persists, etc. Likewise, it is imperative to use and integrate the specific semantics of the business. Finally, the **data lifecycle management** is essential in these environments where large amounts and variety of data are generated, at high speed, so the consumption and processing of data occurs quickly and their outcomes, in turns, are shared by several business processes. Therefore, it is necessary to identify each data and be able to trace its evolution, that is, to know when it appears or is generated in the system, the transformations that it suffers, where it can be found, in what tasks and processes it is involved and its relationship with the strategic business objectives.

Monitoring.

In order to exercise authority, shared decision-making and control required by DG, it is necessary to establish a monitoring process of all the activities involved in the DG program to ensure compliance with the established objectives, policies and standards and ensure that the processes related to information management are correctly executed. In addition, the correct compliance with the agreements (SLA's) established with Cloud service providers must be monitored. This supervision will be based on the definition of the corresponding KPIs that allow to evaluate the fulfilment of the objectives and the creation of the necessary records to be able to carry out a traceability of the DG activities.

4. Conclusions

Enterprise data are now one of the most critical of company assets. Therefore, data must be managed and governed as any other valuable company asset especially in regulated industries and global markets. In addition, the complexity of modern industrial supply chains makes today's data pipelines even more complex having to deal with hundreds of source systems (on-premises and in the cloud) and thousands of sensors and machines, collecting, integrating, cleansing, preparing, relating, protecting and delivering trusted data to be consumed by applications and analytic project users. Thus, defining a simple, maintainable and extendible DG framework for 3GP in I4.0 is highly necessary. Given that the architecture is data-oriented, we consider that the data life cycle is the central element of our proposal, supported on three pillars: an adequate metadata management, data quality and risk management (security and privacy) and all of it governed by policies directed towards company goals. As avenue for future work, we plan to develop models, templates and tools that help companies to apply our framework in real case studies. As a first step we are developing a reference global metadata model and the associated processes for the provision / consume of data in the platform.

Acknowledgements

This work is partially funded by Spanish Government through the research project TIN2017-86520-C3-3-R.

References

[1] Deloitte Consulting LLP, Industry 4.0. Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies, Deloitte, 2015.

[2] I-SCOOP, 'Industry 4.0: The Fourth Industrial Revolution - Guide to Industry 4.0', 2018, pp. 1–58 <https://www.i-scoop.eu/industry-4-0/> [accessed 4 December 2018].

[3] Kevin Kwang, '"Third Platform" Shift Triggers Enterprise Software Evolution', ZDNet, 2013, pp. 1–4 <https://www.zdnet.com/article/third-platform-shift-triggers-enterprise-software-evolution/> [accessed 26 October 2018].
[4] Industrial Internet Consortium, The Industrial Internet of Things Volume G1: Reference Architecture, Industrial Internet Consortium, 2017, pp. 1–58 <https://doi.org/10.1016/j.envsoft.2007.01.004>.
[5] Conor Doherty and others, Building Real-Time Data Pipelines, 2015.
[6] Vijay Khatri and Carol V. Brown, 'Designing Data Governance', Communications of the ACM, 53.1 (2010), 148 <https://doi.org/10.1145/1629175.1629210>.
[7] Ibrahim Alhassan, David Sammon, and Mary Daly, 'Data Governance Activities: An Analysis of the Literature', Journal of Decision Systems, 25. December (2016), 64–75 <https://doi.org/10.1080/12460125.2016.1187397>.
[8] collibra, The Need for Big Data Governance, 2018 <https://mapr.com/resources/need-big-data-governance/assets/the_need_for_big_data_governance_collibra_mapr.pdf>.
[9] Barbara Kitchenham and S Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering_Presentación, Engineering, 2007 <https://doi.org/10.1145/1134285.1134500>.
[10] Deloitte Consulting LLP and others, Data Governance for Next-Generation Platforms, 2018.
[11] Majid Al-ruithe and Elhadj Benkhelifa, 'A Conceptual Framework for Cloud Data Governance-Driven Decision Making', in 2017 International Conference on the Frontiers and Advances in Data Science (FADS) (IEE, 2017), pp. 1–6.
[12] Majid Al-Ruithe, Elhadj Benkhelifa, and Khawar Hameed, 'A Conceptual Framework for Designing Data Governance for Cloud Computing', in Procedia Computer Science (The Author(s), 2016), XCIV, 160–67 <https://doi.org/10.1016/j.procs.2016.08.025>.
[13] Jorge Merino and others, 'A Data Quality in Use Model for Big Data', Future Generation Computer Systems, 63 (2015), 123–30 <https://doi.org/10.1016/j.future.2015.11.024>.
[14] Massimo Felici, Theofrastos Koulouris, and Siani Pearson, 'Accountability for Data Governance in Cloud Ecosystems', in Proceedings of the 2013 IEEE International Conference on Cloud Computing Technology and Science-Volume 2, 2013, pp. 327–32 <https://doi.org/10.1109/CloudCom.2013.157>.
[15] Massimo Felici and Siani Pearson, 'Accountability, Risk and Trust in Cloud Services Towards an Accountability-Based Approach to Risk and Trust Governance', in 2014 IEEE WORLD CONGRESS ON SERVICES (SERVICES), ed. by R Zhang, LJ and Bahsoon, IEEE World Congress on Services, 2014, pp. 105–12 <https://doi.org/10.1109/SERVICES.2014.29>.
[16] Fernando Sáez, 'Aplicación de Cloud Data Lakes En La Industria', 2016, pp. 1–5 <https://www.fernandosaenz.com/aplicacion-de-cloud-data-lakes-en-la-industria/> [accessed 4 June 2018].
[17] M Felici and others, 'Bringing Accountability to the Cloud: Addressing Emerging Threats and Legal Perspectives', Communications in Computer and Information Science, 182 CCIS (2013), 28–40 <https://doi.org/10.1007/978-3-642-41205-9_3>.
[18] Amber Lee Dennis, 'Case Study: Implementing Data Governance in Data Lakes and Big Data', Dataversity, 2017, pp. 1–3 <http://www.dataversity.net/case-study-implementing-data-governance-data-lakes-big-data/> [accessed 28 March 2018].
[19] G Cheng and others, 'Cloud Data Governance Maturity Model', 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2017, 517–20 <https://doi.org/10.1109/ICSESS.2017.8342968>.
[20] Charikleia Paschalidi, 'Data Governance : A Conceptual Framework in Order to Prevent Your Data Lake from Becoming a Data Swamp' (Luleå University of Technology, 2015).
[21] Bitwise, 'Data Governance for Hadoop Enabled Enterprises', 2018.
[22] Majid Al-Ruithe, Siyakha Mthunzi, and Elhadj Benkhelifa, 'Data Governance for Security in IoT & Cloud Converged Environments', in Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2016, pp. 1–8 <https://doi.org/10.1109/AICCSA.2016.7945737>.
[23] Hee Yeong Kim and June-Suh Cho, 'Data Governance Framework for Big Data Implementation with a Case of Korea', in 2017 IEEE International Congress on Big Data (BigData Congress) (IEE, 2017), pp. 384–91 <https://doi.org/10.1109/BigDataCongress.2017.56>.
[24] Philip Russom, Data Lakes: Purposes, Practices, Patterns and Platforms, 2017.
[25] Datameer, 'Datameer Big Data Governance', 2015, 1–8.
[26] Majid Al-Ruithe, 'Development and Evaluation of a Holistic Framework and Maturity Assessment Tools for Data Governance in Cloud Computing Environments' (Staffordshire University, 2018).
[27] David Nuñez, Carmen Fernández-Gago, and Jesús Luna, 'Eliciting Metrics for Accountability of Cloud Systems', COMPUTERS & SECURITY, 62 (2016), 149–64 <https://doi.org/10.1016/j.cose.2016.07.003>.
[28] Alberto Asquer, Governance of Big Data: Perspective and Issues, New Directions in Understanding the Governance of Complex Systems, 2013 <https://doi.org/http://dx.doi.org/10.2139/ssrn.2272608>.
[29] V Tountopoulos and others, 'Interoperability Analysis of Accountable Data Governance in the Cloud', Communications in Computer and Information Science, 470 (2014), 77–88 <https://doi.org/10.1007/978-3-319-12574-9_7>.
[30] Mandy Chessell and others, The Journey Continues. From Data Lake to Data-Driven Organization, ed. by IBM Corporation, 2018.
[31] S Sicari and others, 'Toward Data Governance in the Internet of Things', Studies in Computational Intelligence, 715 (2018), 59–74 <https://doi.org/10.1007/978-3-319-58190-3_4>.
[32] Data Administration Management Association, DAMA-DMBOOK2: Data Management Body of Knowledge, ed. by Deborah Henderson and Susan Earley, 2nd edn (Basking Ridge, New Jersey: Technics Publications, 2017, 2017) <http://catalogo.unican.es/cgi-bin/abnetopac/?TITN=406467>.
[33] Gwen Thomas, 'The DGI Data Governance Framework', The Data Governance Institute, Orlando, FL (USA), 2006, 20 <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+DGI+Data+Governance+Framework#0>.
[34] MDM Institute, 'MDM Institute - Data Governance Definition', 2018 <http://www.tcdii.com/whatIsDataGovernance.html>.
[35] Kristin Weber, Boris Otto, and Hubert Österle, 'One Size Does Not Fit All---A Contingency Approach to Data Governance', Journal of Data and Information Quality, 1.1 (2009), 4:1--4:27 <https://doi.org/10.1145/1515693.1515696>.
[36] Chuck Ballard and others, IBM Information Governance Solutions, ed. by IBM Corporation, 2014 <http://www.redbooks.ibm.com/abstracts/sg248164.html>.