ICIC Express Letters
Volume **13**, Number **9**, September **2019**

ICIC International ©2019 ISSN 1881-803X
pp. 761–771

KEY ASPECTS OF DATA MANAGEMENT FRAMEWORK FOR EARLY ADOPTER: A SYSTEMATIC LITERATURE REVIEW

Wirianto Widjaya[1], Indrajani Sutedja[1] and Armand Wahyudi Hartono[2]

[1]Information Systems Department
School of Information Systems
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ wirianto.widjaya; indrajani }@binus.ac.id

[2]PT. Bank Central Asia Tbk.
Jakarta 10310, Indonesia
armand_hartono@bca.co.id

Received February 2019; accepted May 2019

Abstract. *Data nowadays is a strategic enterprise asset. Data helps organization to get valuable insights, allows itself to be innovative and provides ability to develop its unique competitive advantage. Despite of its strategic role, many organizations neglect the necessity to manage their data properly, mainly due to its broad scope, and complexities. The strategic value of data could not be obtained only by capturing and storing it. Instead, data management initiative is required to develop, execute and supervise plans, policies, programs and processes to deliver, control, protect and enhance the data. The entire data lifecycle process needs to be properly planned and managed in order to ensure data integrity, data quality, data security, and data privacy. This study aims to brief the key aspects of data management framework for early adopters based on Systematic Literature Review (SLR) on selected 25 relevant papers. As the conclusion, the study has identified multiple data management frameworks that are available to provide sufficient guide for data management implementation. Organization could selectively customize the framework implementation based on their specific needs and readiness. Organization also can choose to implement data management framework in stages/incremental manner.*

Keywords: Data management, Data governance, Framework, Big data, Data quality, Data security

1. **Introduction.** Data and information are the most important and strategic business resource in information age. It helps organization to gain competitive advantage over others, to operate faster, in more efficient and more effective manner, and providing support for timely *fact-based* strategic and operational decision [1]. Data is also an asset, an economic resource that can be owned or controlled, holds produce value, and can be converted into money [2]. The world's most valuable resource is no longer oil, but data [3,4]. Many titan size business corporations such as Alphabet (Google's parent company), Amazon, Apple, Facebook and Microsoft yield billion of profits by providing data and information as their key value proposition [3]. Another example is a company named Cambridge Analytica. Despite of being unethical, it has demonstrated the enormous potential of data usage, by doing something that is unthinkable before, which is to help win US presidential runs [5]. The new wave of digital transformation is another source of disruption. Business is rapidly *digitizing*, a term that describes the creation of valuable data and information out of life or business events. It is then followed by *business digitalization*, where business process is reengineered to leverage on the dissemination of data quickly and securely, to get the job done or to solve problem in a faster, more efficient and

DOI: 10.24507/icicel.13.09.761

in the best way possible [5]. Organizations move towards a *digital business* [6] that breaks down boundaries and industry barriers, creates new business opportunity, and destroys old-fashioned and long-successful business model [7]. In order to unleash the potential benefit of data, it needs to be managed properly. Data management implementation then becomes imperative.

This research aims to answer the following main question: "What are the key aspects of data management framework?". The purpose of this study is to identify and comprehend the entire core concept of data management, and provide preliminary essential knowledge to any needy organization as potential early adopters. The research result will help organization to understand the available implementation options, and figure out the very first steps towards the implementation of the data management program based on its strategic goals and readiness. This paper is organized into the following structure: (i) introduction to theoretical foundation of research; (ii) the explanation of research methodology used; (iii) the research results and discussion and (iv) the implication, limitation and conclusions of the research.

2. Theoretical Foundation.

2.1. Data, information, knowledge and wisdom.

Data is the representation of fact or simple observation of the state of world [1,2]. Data represent other than itself, interpret other object and need to be interpreted [2,9]. Data requires a relevant context in order to be meaningful and conventional to understand it and be used as information. Information that is synthesized and contextualized with personal experience is further described as knowledge. When fused with intuition and judgment, knowledge could turn into wisdom, an ability to make accurate decision by applying synthesized and contextualized facts about the world [1].

2.2. Data management.

Data management is a corporate service that executes series of process to develop, execute, and supervise plans, policies, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout its lifecycles [2,13]. Proper data management helps organization to get the strategic benefit by mining wisdom out of it [2,8]. Due to its complexities, data management should be carefully planned and executed. Data management frameworks consistently pointed out the need for organization to customize the implementation based on actual needs and organization readiness [2].

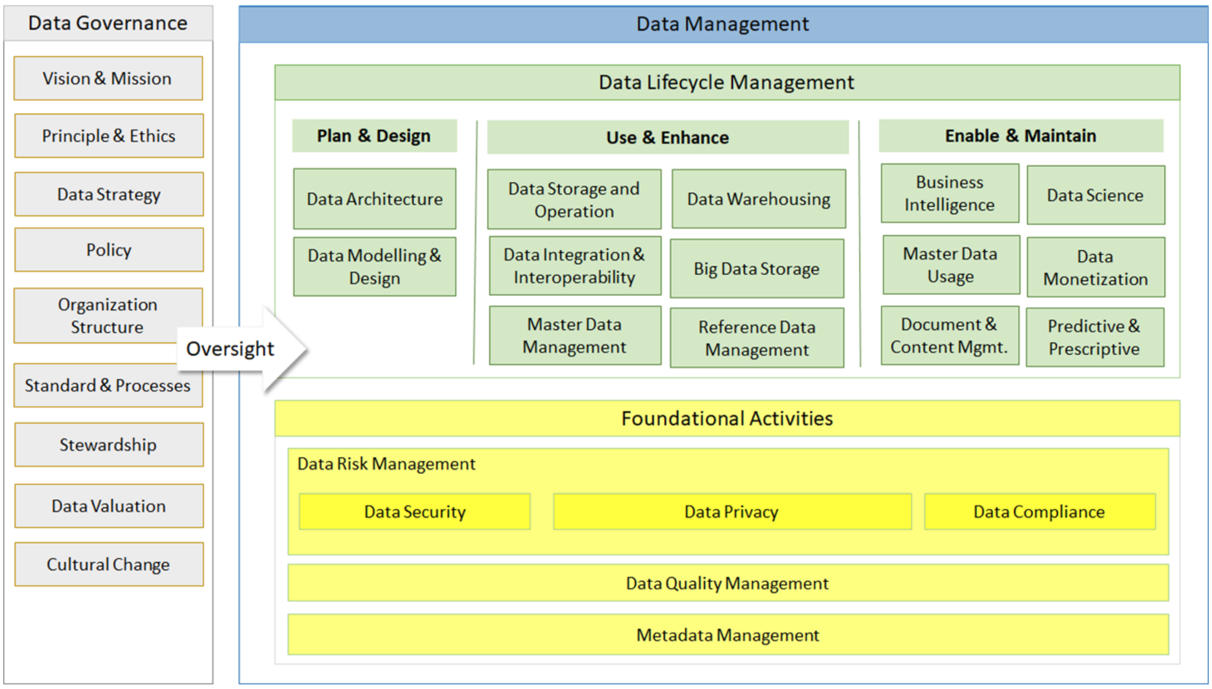
2.3. Data governance.

Data governance is defined as the exercise of authority and control: planning, monitoring, and enforcement, over the management of data as assets [2,10]. It consists of people, processes, and technologies that manage, protect, and use data as an organizational asset [11]. It formulates policies to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions [12]. Data governance guides all data management functions, and ensures all data are managed according to policies and best practices [2]. It comprises systems of formal decision rights and accountabilities for data related process [10]. Data governance is considered as part of data management initiatives [2,10].

3. Methodology.

This study refers to SLR approach as per described by Siddaway [14]. SLR is explained as a systematic study that aims to identify, critically evaluate, and integrate multiple findings and conclusions from various individual studies addressing variety of individual research questions [14]. SLR paper should be able to: (i) preview the progress of existing research towards particular problems; (ii) identity relations, contradictions, gaps and inconsistencies in the literature; (iii) formulate general statements on the overall concept discussed in individual study; (iv) comment on, evaluate, extend or develop a completing theory; (v) provide implications analysis on the practice and policy and (vi) describe the direction and the requirement for future research [15]. The following process step is executed in order to address the research question of this study.

3.1. **Key term elaboration.** To start with, this study searched for conceptual elaboration of "data management" and "data governance" as initial key terms. This study found an **established framework for Data Management**, namely the Data Management Body of Knowledge (**DAMA DMBOK**), that is developed by Data Management International (DAMA International), that aims to advance the concept and practices of data management. DAMA DMBOK suggests that in order to succeed with data management initiatives, organization should cater multiple *knowledge areas* and it supports *environment factors*, as per depicted in Figure 1. To complement DAMA DMBOK framework, this study also reviews another data management framework called Discipline Agile Framework for data management (DA Framework), that is developed by Discipline Agile Consortium (DAC). It aims to promote a pragmatic and streamlined approach to data management [16]. DA Framework discloses several process goals that need to be supported by data management initiatives, which are: (i) improve data quality; (ii) evolve data assets; (iii) ensure data security; (iv) specify data structures; (v) refactor legacy data sources; and (vi) govern data. This study also reviewed the DGI data governance framework from data governance institute in order to gain further elaboration of data governance aspect [10]. As the result, it brings initial understanding of the data management, to be used as search keywords that represent the key aspect of data management.



Sources: DAMA DMBOK by DAMA International

FIGURE 1. Data management overall landscape

3.2. **Study search.** In order to understand the current progress of existing research that is relevant to the research question, this study searches for similar research based on key terms that are identified in previous step. The search is performed in selected sources, including: (i) Google Scholar (https://scholar.google.co.id/); (ii) Science Direct (www.sciencedirect.com); (iii) IEEEXplore Digital Library (http:/ieeexplore.ieee.org); (iv) ACM (http://dl.acm.org/); (v) Springer Link (link.springer.com); and (vi) Research Gate (www.researchgate.net). The key term is used as search keywords, such as: ("Data Management") OR ("Data Governance"); ("Data" AND "Management") AND ("Data"

AND "Architecture"), ("Data" AND "Management") AND (("Data" AND "Modelling") OR ("Data" AND "Design")).

3.3. **Study selection.** Based on study search result, the study tabulates the number of studies that match search keywords as "study found". The title of each listed study is examined to determine its relevancy. When title is not adequate, the abstract section will be further read to decide whether or not a particular study is downloaded for further reading. The number of studies that is downloaded will be counted as "Candidate Studies". All downloaded studies are thoroughly read to check whether a particular study answers and correlates with the research question and later classified as "Selected Study". Total count of selected study for each source will be put into "Selected Study" column for each source. The study selection process prioritized the newest date of publications for the last 5 years. If the count of selected study is more than 3 research per search keywords, the top 3 newest publications will be selected. Table 1 summarizes the result of study selection step.

Table 1. Number of relevant studies found based on identified key terms

No.	Sources	Study found	Candidate studies	Selected studies
1	Science Direct	87	11	2
2	Google Scholar	77	22	10
3	IEEE	22	15	3
4	ACM	13	5	2
5	Springer	3	5	2
6	Research Gate	25	12	5
Total		227	70	24

3.4. **Analysis of selected study.** Each selected study is then analyzed to formulate and compile a general statement, substantive comment, findings or future directions for each key aspect of data management. The general statement for each data management key aspect will be added to result discussion section.

4. **Results and Discussions.**

4.1. **Data management and data governance.** Data management is cross-functional and multi-discipline efforts. It requires collaboration among business and technical units within organization. Due to the complex nature of data management, organization needs to establish well-structured data governance to provide direction, oversight data management activities and coordinate the cross-functional collaboration between various organization unit. Data governance and data management should co-exist in order to ensure each other's effectiveness and success.

4.2. **Data management overall landscape.** Data management has a broad scope. Numerous activities types are included, to ensure that all aspects related to it are well catered. Figure 1 confirmed it. Without proper preparation, it tends to fail. Accordingly, data management implementation will not be easy. Whilst resources are limited, the data management implementation requires prioritization and needs to be conducted in stages. The implementation of data management framework should be customized to meet certain expectation, specific needs and unique focus of each organization.

4.2.1. *Data governance.* Data governance oversights data management to ensure its compliance to direction and policy, and meet the objectives that are set. Data governance organization structure could be centralized or distributed [2,17]. Data governance allows organization to: (i) set up visions and missions; (ii) state principles and ethics; (iii) define data utilization strategy; (iv) develop corporate-wide data policy; (v) set up organization structure; (vi) formalize standards and processes; (vii) regulate and coordinate *data stewardship*; (viii) initiate and organize the measurement of financial value of data; and (ix) lead the related change management activities.

4.2.2. *Data lifecycle management.*

4.2.2.1. *Plan & design.*

4.2.2.1.1. *Data architecture.* Data architecture models the data elements/data structure, relation of one data entity to other, and high level data flows at organization level. It provides the level of abstraction that can be understood by stakeholders, and allow decision to be made accordingly. It helps the organization to standardize the data structure and data content, and decide which system initially produces particular data and how is it going to be changed and used.

4.2.2.1.2. *Data modelling and design.* It is the process of discovering, analyzing, and scoping data requirements for the purpose of presenting and communicating data requirement to respective stakeholder [2]. It serves the following purposes: (i) common vocabulary of data; (ii) documentation of organization's data on system; (iii) communication for data architecture and requirement; and (iv) the starting point for application development or enhancement. There are several data modelling schemes, namely, relational, dimensional, object oriented, fact-based, and NoSQL Selection is done based on the technology used by project to meet specific organization's requirement [18].

4.2.2.2. *Use & enhance.*

4.2.2.2.1. *Data storage and operation.* It defines and executes database management activities throughout data lifecycle, from initial implementation, obtaining data, backup, and purge data [2]. It is designed to ensure: (i) the availability of data; (ii) the integrity and security of data assets within storage; and (iii) to achieve optimal level of data transactions performance.

4.2.2.2.2. *Data integration and interoperability.* It identifies and defines processes to move data between data stores, application and organizations, including data interfacing requirement between systems/application within and inter organizations. The result is used to identify the overall functional and technical requirement to ensure data quality, data integrity, data security, and to meet other specific requirements such as web semantic requirement [2,19]. Based on it, organization selects a standard tool and technology for data integration and interoperability.

4.2.2.2.3. *Reference and master data management.* Reference and master data is shared data that is used and maintained across business areas, processes and systems, such as country codes, customer data, and governmental tax codes. It needs to be consistently managed by implementing Master Data Management (MDM). MDM resolves data duplication, inaccuracy, and inconsistency, and reduces the cost of data integration and consolidation [2,20]. MDM also caters: (i) master data sharing; (ii) provision of single version of truth and authoritative source of master data; and (iii) to lower the complexities and cost of data integration and consolidation. MDM is a broad and complex activity, where another study defines MDM maturity model with 13 focus area and 65 capabilities [38].

4.2.2.2.4. *Data warehousing.* In order to allow effective analysis within organization, data from various sources needs to be consolidated. Data warehousing is the process of planning, implementation, and control of data consolidation from various data sources into single location and a common data model. It comprises the build and maintenance of technical environment, and the execution of processes needed to support business intelligence activities [2].

4.2.2.2.5. *Big data storage.* Big data serves the requirement of collecting various types of data to find new facts and insights. Big data is characterized by 10V category: (i) massive volume; (ii) high speed velocity; (iii) variety types and format; (iv) veracity, the messiness of data; (v) variability or multiple meaning; (vi) visual representation; (vii) data value; (viii) validity, the correctness and accuracy of data; (ix) distributed and heterogeneous data source venue; and (x) vagueness, the lack of certainty [21]. Big data is valuable to: (i) discover data relationship; (ii) support the massive data consolidation; (iii) enable the discovery and analysis of new influencing factor for business [2]; and (iv) reveal the potentially-hidden tacit knowledge inside data [22]. Big data is a must-have initiative to complement data warehouse.

4.2.2.3. *Enable & maintain.*

4.2.2.3.1. *Business intelligence.* Business Intelligence (BI) is a combination of architecture, tools, databases, application, methodologies and data analysis to gain understanding on organizational activities and opportunities [2,23]. It includes: (i) reporting; (ii) dashboard; (iii) ad hoc queries; (iv) search and lookups; (v) OLAP; (vi) interactive visualization; (vii) scorecards; (viii) predictive modelling; (ix) data mining; (x) data mining workbench; (xi) column based DBMS; (xii) in memory database; and (xiii) real-time decision tools [24]. BI visualizes insight and data pattern in the most optimal, intuitive and understanding-oriented way, and is served as an essential gateway to harvest knowledge and wisdom out of organization's data.

4.2.2.3.2. *Master data usage.* Organization should facilitate reference and master data sharing to allow consistent usage for all data stores, application and systems [39]. There are 3 approaches, which are: (i) registry, sharing from originated system; (ii) transaction hub, sharing from centralized locations; and (iii) hybrid approach, managed in originated system and then centrally share from a single location [2].

4.2.2.3.3. *Document & content management.* Besides structured data, organization also produces and collects various unstructured data in the form of documents, such as protocols, procedures, methods and specification, or records, such as contract, memos, and reports. In many of the cases, volume of unstructured data outnumbers structured data [37]. The planning, implementation and control activities of data and information of unstructured data are called document and content management [2]. It allows to: (i) comply with legal/compliance requirement and customer expectations; (ii) effectively and securely store, retrieve and use documents and records; and (iii) provide integrated analysis capabilities of structured and unstructured data. Enterprise Content Management (ECM) is the key technology for document and content lifecycle management, access right management and semantic search capabilities. Currently, there is no solution to partially display a single document that contains sensitive data of two or more parties.

4.2.2.3.4. *Data science.* Data science is the application of interdisciplinary scientific field to extract knowledge and insight from structured and unstructured data to answer specific business questions and to support strategic business decision [25]. It performs automated analysis of data that unifies the use of principles, processes, techniques, scientific method and technologies to reveal hidden pattern, predicts the future data behavior and outcome and identifies phenomenon within data. Data science implementation is supported by big data storage for consolidated data provision and business intelligence for data visualization.

4.2.2.3.5. *Data monetization.* Data monetization is a direct way to get value from data by selling it to other party [2], or by providing value added services such as customer micro-segmentation, location based marketing, customer sentiment analysis, and price comparison [27]. Data become the primary product to generate revenue stream for organization.

4.2.2.3.6. *Predictive & prescriptive analytics.* Predictive and prescriptive analytic model is part of data science application. Predictive analytic allows organization to predict the future using a model that is based on pattern found inside the data. Prescriptive analytic applied mathematical model that yields foresight to "prescribes" optimal behaviors and action to best meet organization's goals [28,29,36]. Both methods are used to provide the best course of actions to anticipate the future or to measure and improve performance.

4.2.3. *Foundational activities of data management.* Foundational activities are compulsory common services and basic structures to ensure proper and risk-mitigation-oriented execution of data lifecycle activities.

4.2.3.1. *Data security.* Data security is the planning, development and execution of policies and procedures to ensure the security of data. It provides user authentication and authorization of access, as well as related auditing data [2]. Data security should ensure the overall: (i) integrity; (ii) confidentiality; (iii) availability; and (iv) accountability [31]. Its goals are to: (i) enable appropriate access and prevent inappropriate access to data; (ii) comply with regulations and policies; (iii) enforce data privacy and confidentiality protection. Related aspects of data securities are: (i) system vulnerability; (ii) threats; (iii) data loss or unauthorized access risk, (iv) the protection of data integrity; and (v) data encryption, particularly in cloud environment [30].

4.2.3.2. *Data privacy.* Organization which collects personal data needs to prevent personal data errors or misused that might impact individual's well-being or even impact personal safety and interest [2]. There are related issues such as: (i) no awareness on the type of personal data and its location; (ii) no customer consent for the use of personal data; and (iii) share of personal data with other organization that has no proper personal data handling [32].

Data privacy implementation includes: (i) guiding principle on personal data usage and handling; (ii) the risk assessment if there are open accesses to personal data; (iii) sets of handling practice and procedure; and (iv) review, audit and assurance of data privacy protection [2]. Enterprise Privacy Practice (E-P3P) is one example of data privacy management framework consisting of related technology and processes [32].

4.2.3.3. *Data compliance.* It is very critical for organization to meet data-related regulatory compliance requirement based on laws and regulation applied to its area of industry [2]. To start with it, organization needs to: (i) assess the compliance level of data related processes and practices; (ii) identify data privacy vulnerability; and (iii) assess the impact of compliance incidents [33]. To ensure data compliance, organization needs to: (i) identify data regulatory requirement, (ii) integrate data compliance requirement to data quality management processes; (iii) perform issue management on data compliance incidents; and (iv) perform hands-on observation, audit, and correction with regard to data compliance management [2].

4.2.3.4. *Data Quality Management (DQM).* In order to be able to bring the intended and expected value, data need to be trustworthy, reliable, and constant in high quality state [40]. Data quality dimensions are: (i) availability; (ii) usability; (iii) reliability; (iv) relevance; and (v) presentation quality [34]. DQM is also important for data integration between organizations [42]. It aims to: (i) increase the value of data; (ii) reduce risk and costs; (iii) improve organization efficiency and productivity; and (iv) protect and enhance organization's reputation. DQM activities include: (i) development of governed approach to make data fit for intended usage purpose; (ii) define standards, requirements and specification for data quality controls; (iii) define and implement processes to measure, monitor and report data quality; and (iv) identify and advocate opportunities to improve data quality through processes or system implementation [2].

Table 2. Summary of key aspects and DAMA DMBOK knowledge areas mappings

No.	Group/explained key aspect	DMBOK2 knowledge area
1	4.2.1 Data governance [2,17]	Data governance [2]
	4.2.2 Data lifecycle management	
	4.2.2.1 Plan & design	
2	Data architecture [2]	Data architecture [2]
3	Data modelling and design [2,18]	Data modelling & design [2]
	4.2.2.2 Use & enhance	
4	Data storage & operation [2]	Data storage & operation [2]
5	Data integration and interoperability [2,19]	Data integration and interoperability [2]
6	Reference and master data management [2,20,38]	Reference and master data management [2]
7	Data warehousing [2]	Data warehouse and business intelligence [2]
8	Big data storage [2,21,22]	Big data and data science [2]
	4.2.2.3 Enable & maintain	
9	Business intelligence [2,23,24]	Data warehouse and business intelligence [2]
10	Master data usage [2,23]	Reference and master data management [2]
11	Document and content management [2,37]	Document and content management [2]
12	Data science [2,25,26,41]	Big data and data science [2]
13	Data monetization [2,27]	Data governance [2]
14	Predictive & prescriptive analytics [2,28,29,36]	Big data and data science [2]
	4.2.3 Foundational activities of data management	
15	Data security [2,30,31]	Data security [2]
16	Data privacy [2,32]	Data security [2]
17	Data compliance [2,33]	Data security [2]
18	Data quality management [2,34,40,42]	Data quality [2]
19	Metadata management [2,35]	Metadata [2]

4.2.3.5. *Metadata management.* Metadata is all kind of data and information that describes other data, including: (i) business context of data, such as data definition, concepts its represents, and data privacy level; (ii) technical information such as data elements, database, data models, application and systems; and (iii) operational information such as data update status, and last error found in data [2]. Metadata serves as data catalog, for both structured and unstructured data, that help data consumer to find and understand it quickly [35].

Metadata management is the series of activities of planning, implementation, and controls to enable high quality and integrated metadata. The primary goals are to: (i) provide understanding on data and related business terms/glossaries; (ii) collect and integrate metadata from various sources, (iii) provide standardized and secured way to access metadata; and (iv) ensure metadata quality and security.

5. **Implications, Limitations, and Conclusions.** This paper serves as complete and quick reference to overall aspect of data management initiatives for organization and scholars who are interested in data management topic. The core concept and essential attribute of each data management aspect is explained to bridge the next effort for detail comprehension and implementation of each respective knowledge area. The number of papers that is reviewed in this study is limited due to restricted access to paper database. This study prioritized papers that are published within the last 5 years. This study found 19 distinctive key aspects of data management due to the broad-scope and complex nature of the initiatives. Based on defined business priorities, identified stakeholder requirements, and readiness, organization should plan the data management implementation roadmap. The first priority should be given to basic data operation management activities: (i) data modelling and design; (ii) data integration and interoperability; (iii) data storage and operations; and (iv) data security. In this way, organization shall benefit immediately. As the next step, organization could choose to implement selective part of data governance initiatives to provide direction and to develop enterprise strategy and detail planning for data management implementation throughout the organization that is aligned with organization strategic objectives. Further study is still required to understand how each data management key aspect is selectively implemented to fit into specific requirement and condition of different organizations from different industries.

Acknowledgment. This work is partially supported by the respective reviewers. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] K. E. Pearlson, C. S. Saunders and D. F. Galletta, *Managing and Using Information Systems, Binder Ready Version: A Strategic Approach*, John Wiley & Sons, 2016.
[2] P. Cupoli, S. Earley and D. Henderson, *DAMA-DMBOK2 Framework*, 2014.
[3] Economist Team, *The World's Most Valuable Resource Is No Longer Oil, But Data*, The Economist, New York, NY, USA, https://www.economist.com, accessed on December 12, 2018.
[4] J. Toonders, Data is the new oil of the digital economy, *WIRED*, http://www.wired.com, accessed on December 12, 2018.
[5] C. Cadwalladr and E. Graham-Harrison, *The Cambridge Analytica Files. I Made Steve Bannon's Psyhological Warfare Tool: Meet the Data War Whistleblower*, https://www.theguardian.com, accessed on December 12, 2018.
[6] J. Bloomberg, *Digitization, Digitalization, and Digital Transformation: Confuse Them at Your Peril*, https://www.forbes.com, accessed on December 12, 2018.
[7] P. Weill and S. L. Woerner, Thriving in an increasingly digital ecosystem, *MIT Sloan Management Review*, vol.56, no.4, p.27, 2015.
[8] M. Fleckenstein and L. Fellows, *Modern Data Strategy*, Springer International Publishing, 2018.

[9] L. Sebastian-Coleman, *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework*, Newness, 2012.

[10] G. Thomas, *The DGI Data Governance Framework*, The Data Governance Institute, Orlando, FL, USA, http://www.datagovernance.com, accessed on December 12, 2018.

[11] H. J. Watson and M. McGivern, Getting started with business-driven data governance, *Business Intelligence Journal*, vol.21, no.1, pp.4-7, 2016.

[12] S. Soares, *Data Governance Tools: Evaluation Criteria, Big Data Governance, and Alignment with Enterprise Data Management*, Mc Press, 2015.

[13] K. Gordon, *November. Principles of Data Management: Facilitating Information Sharing Second Edition*, BCS, 2013.

[14] A. Siddaway, *What Is a Systematic Literature Review and How Do I Do One?*, https://www.semanticscholar.org/paper/WHAT-IS-A-SYSTEMATIC-LITERATURE-REVIEW-AND-HOW-DO-I-Siddaway/22142c9cb17b4baab118767e497c93806d741461#citing-papers, 2014.

[15] R. F. Baumeister, Writing a literature review, in *The Portable Mentor*, Springer, New York, NY, 2013.

[16] *The Disciplined Agile Consortium*, The Disciplined Agile (DA) Framework, http://www.disciplinedagiledelivery.com/agility-at-scale/data-management/, accessed on November 25, 2018.

[17] F. Nadeem, A taxonomy of data management models in distributed and grid environments, *International Journal of Information Technology and Computer Science*, vol.8, no.3, pp.19-32, 2016.

[18] V. I. Nnebedum, A. U. Kamalu and J. N. Dike, From data management to data engineering, *International Journal of Computer Applications*, vol.81, no.11, pp.36-39, 2013.

[19] S. K. Bansal, Towards a semantic Extract-Transform-Load (ETL) framework for big data integration, *2014 IEEE International Congress on Big Data (Big Data Congress)*, pp.522-529, 2014.

[20] F. Haneem, R. Ali, N. Kama and S. Basri, Resolving data duplication, inaccuracy and inconsistency issues using Master Data Management, *IEEE 2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp.1-6, 2017.

[21] S. Kaur and W. Singh, Systematic mapping study of big data mining tools and techniques, *IEEE the 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp.1-7, 2017.

[22] K. Dalkir, *Knowledge Management in Theory and Practice*, Routledge, 2013.

[23] R. Sharda, D. Delen, E. Turban, J. Aronson and T. P. Liang, *Businesss Intelligence and Analytics: Systems for Decision Support*, Prentice Hall, London, 2014.

[24] H. Chen, R. H. Chiang and V. C. Storey, Business intelligence and analytics: From big data to big impact, *MIS Quarterly*, pp.1165-1188, 2012.

[25] V. Dhar, Data science and prediction, *Communications of the ACM*, vol.56, no.12, pp.64-73, 2013.

[26] J. Leek, The key word in 'data science' is not data, it is science, *Simply Statistics*, 2013.

[27] J. M. Cavanillas, E. Curry and W. Wahlster, *New Horizons for a Data-Driven Economy*, SpringerOpen, 2015.

[28] T. H. Davenport, Analytics 3.0, *Harvard Business Review*, vol.91, no.12, 2013.

[29] W. L. Winston and J. B. Goldberg, *Operations Research: Applications and Algorithms*, Thomson Brooks/Cole, Belmont, 2004.

[30] D. C. Kant and Y. Sharma, Enhanced security architecture for cloud data security, *International Journal of Advanced Research in Computer Science and Software Engineering*, vol.3, no.5, 2013.

[31] Y. Demchenko, C. De Laat and P. Membrey, Defining architecture components of the big data ecosystem, *IEEE 2014 International Conference on Collaboration Technologies and Systems (CTS)*, pp.104-112, 2014.

[32] G. Karjoth, M. Schunter and M. Waidner, Platform for enterprise privacy practices: Privacy-enabled management of customer data, in *International Workshop on Privacy Enhancing Technologies*, Springer, Berlin, Heidelberg, 2002.

[33] E. Gelbstein, Quantifying information risk and security, *ISACA Journal*, pp.433-438, 2013.

[34] L. Cai and Y. Zhu, The challenges of data quality and data quality assessment in the big data era, *Data Science Journal*, vol.14, 2015.

[35] M. Varade and V. Jethani, Distributed metadata management scheme in HDFS, *International Journal of Scientific and Research Publications*, vol.3, no.5, pp.2250-3153, 2013.

[36] S. Srinivas and A. R. Ravindran, Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework, *Expert Systems with Applications*, vol.102, pp.245-261, 2018.

[37] S. Laumer, D. Beimborn, C. Maier and C. Weinert, Enterprise content management, *Business & Information Systems Engineering*, vol.5, no.6, pp.449-452, 2013.

[38] M. Spruit and K. Pietzka, MD3M: The master data management maturity model, *Computers in Human Behavior*, vol.51, pp.1068-1076, 2015.
[39] A. Reichert, B. Otto and H. Österle, A reference process model for master data management, *Wirtschaftsinformatik*, p.52, 2013.
[40] K. Weber, B. Otto and H. Österle, One size does not fit all – A contingency approach to data governance, *Journal of Data and Information Quality (JDIQ)*, vol.1, no.1, p.4, 2009.
[41] F. Provost and T. Fawcett, Data science and its relationship to big data and data-driven decision making, *Big Data*, vol.1, no.1, pp.51-59, 2013.
[42] B. Lauro and R. Traverso, Data fitness for integration, *2018 European Conference on Quality in Official Statistics*, Krakow, 2018.