DATA-DRIVEN ENTERPRISE ARCHITECTURE AND THE TOGAF ADM PHASES

Carike Kearny, Aurona Gerber and Alta van der Merwe
Department of Informatics, University of Pretoria, Pretoria, South Africa
carike670@gmail.com, aurona.gerber@up.ac.za, alta@up.ac.za

***Abstract* - This paper investigates how Data as a disruptive technology could be integrated into TOGAF. Given the recent attention of Big Data and Data Science as disruptors, this paper investigates what the impact on the enterprise could be and how Enterprise Architecture (EA) should accommodate data to enable data-driven EA. There is no model currently available that investigates how Big Data can be incorporated into data-driven EA solutions. This study specifically focuses on how the TOGAF ADM could support a data-driven enterprise. Through document analysis and a systematic literature review, a specific adaption of the TOGAF ADM is proposed that indicates the influence that Data and Big Data has on each phase within the ADM.**

I. INTRODUCTION

Data plays an important role within every organization, especially since Big Data are regarded as one of the technology disruptors of the next decade [1]. According to CSC the Big Data Universe is exploding and will continue to do so well into the next decade [2]. However, the challenge of how to optimally and effectively manage large amounts of data as well as identify new ways of analysing this data to unlock information and opportunities remains. This scenario gives rise to the 'Big data' problem, which could be defined as: "The practice of collecting complex data sets so large that it becomes difficult to analyse and interpret manually or using on-hand data management applications" [3]–[5].

In order for enterprises to prepare and integrate data into their strategy and structure, the incorporation of data into the Enterprise Architecture (EA) of the organisation should be investigated. Data science, generally defined as the *science of data*, is fast becoming relevant for any organizational strategy given the envisaged advantages of Big Data as phenomenon [6], [7]. To remain competitive, organizations need to start thinking of ways to integrate data into everyday operations, processes, strategy and technology.

Within the EA domain, TOGAF is one of the most adopted enterprise architecture (EA) frameworks. TOGAF specifically focuses on EA as a mechanism to provide business support and ensure business-IT alignment, as well as to provide essential technology and process structure for an IT strategy in any enterprise. Within TOGAF, the ADM is the process proposed for the implementation of EA and this study specifically focused on the TOGAF ADM [8]. Since EA is regarded as the most significant means to align strategy, processes and IT within an organisation, Data and Big Data should be incorporated into the EA of any organisation. It is thus useful to investigate how Data can be incorporated into EA execution. The TOGAF ADM is regarded as a standard process for ES implementation [9] and Figure 1 illustrates the TOGAF ADM Cycle. Within this study, the TOGAF ADM cycle is the main focus and the integration of Data within each phase is investigated. The contribution of the study is a Data-driven EA perspective given the TOGAF ADM phases [10].

The remainder of this paper is structured as follows: Section II provides background on EA, TOGAF and the TOGAF ADM. Section III introduces Data and Big Data as disruptors. Section IV presents the research approach as well as the findings, Section V interprets the results and Section VI concludes.

II. ENTERPRISE ARCHITECTURE

In order to understand the term enterprise architecture (EA), it is firstly useful to define what is regarded as an enterprise is and furthermore, what is meant by *architecture*. An enterprise is any organization or group of organizations, which share common objectives and that work together to achieve common goals. Enterprise architecture (EA) is concerned with the design of enterprises and according to the Open Group's TOGAF, an architecture is a description or detailed plan of an enterprise at component level that guides the implementation thereof [11]. EA could be defined as a conceptual blueprint that defines the structure and operations of an organization [12], [13]. The current and future objectives of an organization is something that EA focuses on and EA in general aims to assist an organization to effectively reach their strategic objectives.

According to the Open Group it is necessary to develop an EA to provide business support as well as to provide essential technology and process structure for an IT strategy [8]. An EA describes the manner in which an organization will function, the relationships and structure of the organization, its business models as well as the way technology and information systems will support the organizational goals and objectives [11]. EA provides documentation and information that supports planning and decision-making at various levels of scope and detail, which means it should provide major benefits for management. EA provides a strategic context for the growth and change of the IT system in response to the shifting demand of the business environment. In short, Enterprise architecture aligns IT with business. An effective EA can produce many technical and business benefits like more efficient IT operations for instance, reduced IT infrastructure complexity and faster procurement, only to name a few [14], [15].

EA frameworks can be used in any organization to develop an EA because an EA framework is a foundational structure that defines how to produce and use an enterprise architecture implementation [15]. An EA framework in general describes a method and a list of recommended standards for designing and implementing an EA within an organization using a series of phases. Several different EA frameworks exist that are designed by different experts for the needs of different organizations and they thus focus on different facets of an EA implementation [16]. In general it is accepted that the use of an EA framework will increase the effectiveness of an architecture development and will ensure thorough coverage and understanding of the proposed solution. The Open Group has been active in the development of EA standards for several years and TOGAF is the result of an

978-1-5090-1897-0/16/$31.00 ©2016 IEEE

initiative between the Open Group and several industry role players and leaders. As stated, TOGAF is the focus of this study.

A. TOGAF

TOGAF focuses on EA specification and implementation, including enterprise architecture methodology. This study will focus on TOGAF and the TOGAF ADM as specified by the open Group [8], [10].

TOGAF was selected as it is regarded as a proven enterprise architecture methodology and framework that has been adopted by many leading organizations for the purpose of significantly increasing business productivity and business-IT alignment. The TOGAF EA is considered to be one of the most prominent and reliable standards that ensures consistent methods, standards and communication among various EA specialists. Greater industry credibility, career opportunities and job satisfaction is some of the perks that TOGAF claims to provide for those professionals that are fluent in the TOGAF standards. TOGAF could also assist with utilizing resources more effectively and efficiently and can thus ultimately produce an increased return on investment [8].

TOGAF deals with the four architecture domains commonly accepted in enterprise architecture. These domains are:

- Business Architecture – Describes the fundamental business processes, organization, governance and the business strategy.
- Data Architecture – The structure of an organization's physical and logical data management resources and data assets is described in this domain.
- Application Architecture – A blueprint of the individual applications is provided that contains the interactions among applications as well as the relationships they have with the organization's core business processes.
- Technology Architecture – A description of the hardware and software capabilities required to support the deployment of different services is provided. This includes networks, IT infrastructure, processing, communications, standards, etc.

As part of the TOGAF framework, TOGAF provides a method that describes how to develop enterprise architecture. The TOGAF Architecture Development Method (ADM) is a repeatable method for developing enterprise architectures [10].

Figure 1 illustrates the nine phases of the TOGAF ADM EA method. The first phases of the ADM specify and define the EA design of an enterprise, whilst the latter phases (from Phase E onwards), are primarily concerned with EA implementation. For the purposes of this paper, the focus will be on the EA definition phases namely: the Preliminary phase, the Architecture Vision (A), the Business Architecture (B), the Information Systems Architecture (C) and the Technology Architecture (D) phases.

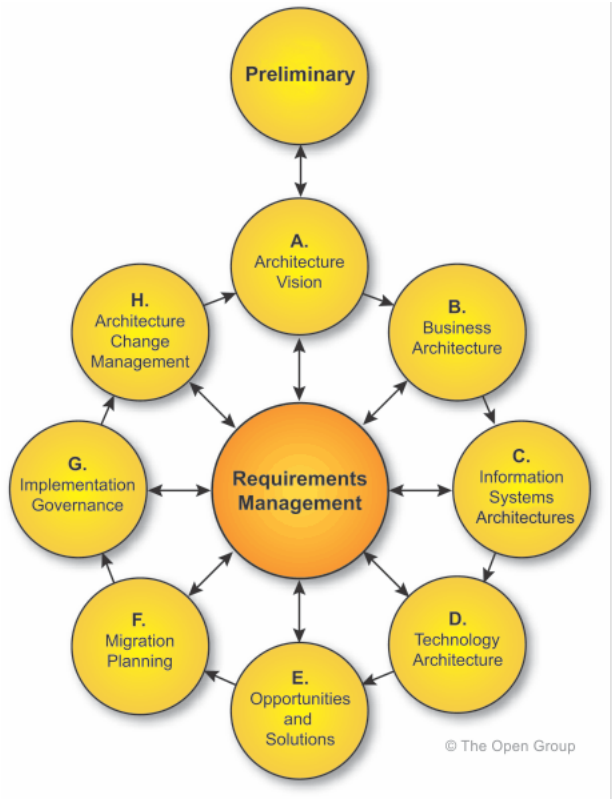

Figure 1: The TOGAF ADM Cycle [10].

B. The TOGAF ADM

The purpose of the TOGAF ADM as previously stated is to help organizations develop an EA suitable to that specific enterprise that encapsulates the organization's strategy and the incorporation of this strategy into all the aspects of the organization namely the vision, the business architecture, the information systems architecture and the technology architecture [10]. A phase represents a component of business, architectural or IT capability that work in conjunction with other phases to provide solutions and implementations. The ADM process forms the core of any TOGAF implementation and makes it possible to establish an architecture framework, develop architecture content as well as implement the EA in the organization. The ADM thus also assists organizations with the transitioning from the old (as-is) architecture to the new (to-be) enterprise architecture. A short description of the first five phases is provided below:

- *Preliminary Phase*
  This phase includes the initiation and preparation activities required to meet the business requirements for new enterprise architecture, the definition of an organization's architecture framework as well as definitions of principles. This phase is about answering the Where, What, Why, Who and How questions for an organizations architecture. The main aspects are to define the architecture, the requirements, the architecture principles, the framework to be used and the relationships between the management frameworks.

- *Architecture Vision*
  Describes the initial stage of the ADM, which includes:
  -Outline the scope
  -Identify stakeholders
  -Acquire approvals
  -Architecture vision is constructed
- *Business Architecture*
  Business Architecture development is defined to support an approved architecture vision.
- *Information Systems Architecture*
  Defines the IS architectures of an architecture project, comprising the development of Application and Data architectures.
- *Technology Architecture*
  The development of Technology Architecture is defined for a specific architect project.

The first phases of the TOGAF ADM is primarily concerned with the design and definition of the (as-is and to-be) EA of an organization. When considering any technology disruptor such as Big Data, we propose that the initial phases of the ADM will have to consider the disruptive technology as part of the EA definition and specification. The next section presents the core concepts of data, Big Data and data science.

III. DATA, BIG DATA AND DATA SCIENCE

Data is a core phenomenon of any information system and consists of the digitized reality that it represents. Data can therefore represent many different things, have many different formats and can be used in many different ways. Data is considered fundamentally as digitized raw facts gathered for analysis to derive certain information from it to be used for a specific purpose [17].

In the 21st century data is regarded as a technology disruptor because it is growing at a tremendous rate [1]. It is becoming harder and harder for organizations to store, curate, manage, manipulate as well as process data into any useful information due to the sheer volume and diversity of the available data. It is estimated that the volume of data is growing 40% per year according to the estimations of The McKinsey Global Institute [18]. Traditional enterprise data, machine-generated data and social data all contributes to this massive data volume that modern enterprises need to deal with. It is said that much of the current economic activity, growth and innovation would not have been possible without data.

Big Data is a term widely used but for the purpose of this study it is defined to refer to data or datasets which size is too vast to be captured, stored, analysed and managed by traditional database software tools. Big data can be seen as large stores of data that can be captured, combined, communicated, stored and analysed. Big Data can be characterised by the following 'V's [7], [19], [20]:

- *Volume:* The data is generated in very large capacities often running into the Petabytes (e.g. machine-generated data). Humans, machines and human computer interactions like social media generate data.
- *Variety:* The data formats change very frequently. Structured (databases, spreadsheets) and unstructured (emails, photos, videos, etc.) data is generated, which in turns produces problems for storage, analysis and mining of data.
- *Velocity:* The frequency with which the data is generated is very high (e.g. social media). Streaming data must be analysed within an acceptable time. If organizations are able to handle the speed by which real-time data is generated, better and informed strategic decisions can be made to yield an increase in competitive advantage.
- *Veracity:* Not all data is of value. The challenge is to identify amongst the large quantities of data what is of value to the enterprise and what is not. The data being used to make decisions in an organization is not always trustworthy or as relevant to the problem at hand. Data cleansing and preparation is of immense importance in organizations.

Associated with discussion on Big Data, the term Data Science also emerged. Data science could be defined as an all-encompassing multi-disciplinary science with the main objective the development of value or products from available data [21]. This definition is supported by Stanton who claims that data science refers to an "emerging area of work concerned with the collection, preparation, analysis, visualization, management and preservation of large collections of data" [22]. Data science thus incorporated Data, Big Data and places an emphasis on making sense of Data and Big Data and in the process produce data products.

C. *The Importance of Data Science in organizations*

In the business world as we know it, strategy, including having a competitive edge and useful business information, are major role players in the success of an organization. Processing, analysing and integrating available data could produce very useful and insightful information to the business, and more importantly, could provide the competitive edge necessary for the business to survive in an increasingly competitive environment. The data-derived information could be used by management for strategic decision-making and can in turn lead to enhanced productivity, a competitive edge and greater innovation, all of which can have a major impact on the organization's success [7].

Data science can add value to organizations in many different ways:

- *Provide transparency*. Making data more easily accessible in a timely fashion can significantly reduce search and processing time.
- *Enables organizations to discover needs, expose inconsistencies and improve performance*. As more transactional data is created and stored, the more precise and detailed performance information can be derived from it. This data can be used to identify and better understand variability or inconsistencies in performance, which means that managers can pinpoint the root causes and manage performance to higher levels.
- *Marketing specific to a target market*. Organizations can now analyse specific

segments of data to customize products and services to cater for those specific needs.
- *Replacing/supporting human decision making.* Sophisticated analytics can reduce risks, improve decision making and provide insightful information that would have otherwise been unknown.
- *Innovating new business models, products and services*. Organizations or manufacturers can use the data obtained from the use of actual products by consumers to improve the standard or quality of the next generation of products.

The importance of data to organizations proves the significance of processing, manageing and incorporating data into an organizations strategy and structure, hence the need to investigate how data impacts on the EA of an orgaization.

D. Data-driven Organizations

In this section some characteristics of typical data-driven enterprises are presented. Generally such data-driven enterprises enjoy advantages over their competitors because of the insights provided by data management and analytics and could, for instance, enhance marketing strategies and planning that involves customer insight. The organizational nature of a data-driven enterprise involves the integration of a data-driven strategy and structure that will influence the enterprise architecture. Data-driven organizations need to consider a wide variety of data that is often widely accessible. Data is everywhere and not only stored on a main server but also off-site or in the Cloud. Generally, insight into the data should receive priority over technology used, thus data should be organized, managed and results should be easy to interpret. Data flows are throughout an organization, up, down as well as side to side in data-driven organizations. Everyone in the organization that make decisions also uses data, it is not only used to inform senior decision makers.

IV. SYSTEMATIC REVIEW

As stated, this study investigated what the impact of the integration of data into an organization would be on the execution of the TOGAF ADM. The study was executed using document analysis and a systematic literature review [23]. A systematic review of the literature was performed to ultimately produce an adapted version of the TOGAF ADM.

The focus was set on data-driven organizations and how the organization's vision and architecture is influenced accordingly. Big data is a disrupter in organizations and due to this it is an important factor to consider. With the maximum use of data as part of the business strategy, the TOGAF ADM is influenced several ways. There was only a high level focus on the first five phases of the ADM, namely Preliminary phase, Architecture Vision, Business Architecture, Information Systems Architecture and Technical Architecture. The outputs and goals of each process was addressed to determine how big data will impact each. A table for each phase of the TOGAF ADM was created to show how intensely big data (the 4 V's) influence the objectives and goals of each phase. At the end of the systematic review a table was created to show the summarized result set that indicates where data has a high, medium or low influence on the processes within the TOGAF ADM.

This systematic literature review process conducted consisted of three stages: Planning, Execution and Result analysis [23].

In the planning phase the research objectives was defined as well as the manner in which the review was executed. The question focus, question quality, source selection and source selection criteria was discussed [23].

The execution phase involved the extracting of the information needed from the defined sources to make an applicable conclusion. A table was compiled for each phase of the TOGAF ADM. The objectives of each phase was mapped against the 4 V's of Big Data to illustrate how each big data characteristic will influence the specific objectives within each ADM phase. A score was given for each objective. The higher the score, the higher the impact of big data on the objective.

The result analysis phase took place after the systematic review execution. The results has been presented in a logical format for the purpose of answering the research question as is presented in Table 1: Result Summary.

Based on the data obtained in the System Review Execution phase, a table was formulated as a summarization of the findings. The following table gives an indication of the extent to which the characteristics of Big Data will influence the certain processes in the TOGAF ADM framework.

Table 1: Result Summary

Process / Characteristic	Preliminary phase	Phase A: Architecture Vision	Phase B: Business Architecture	Phase C: Information Systems Architecture: Data Architecture	Phase C: Information Systems Architecture: Application Architecture	Phase D: Technology Architecture
Volume	72.9%	45.4%	72.4%	100%	100%	97.6%
Variety	46.2%	40%	72.4%	100%	100%	92.9%
Veracity	47.5%	40.8%	71.4%	100%	100%	92.9%
Velocity	62.1%	41.2%	71.4%	100%	100%	97.6%
Total Impact	57.2%	41.9%	71.9%	100%	100%	95.3%

V. INTERPRETATION OF RESULTS

As illustrated in Table 1, data will influence the first five phases of the TOGAF ADM. The Data Architecture, Application Architecture and Technology Architecture will be fully impacted and this is where the biggest challenges for data-driven enterprises lie. Organizations should make a concerted effort to focus on data science mechanisms and the processing thereof to ensure competitive advantage.

An adapted TOGAF ADM model, as derived from Table 1, was created to illustrate the degree of how the 4 V's of Big Data will influence each phase of the ADM as can be seen in Figure 2 below. The first five phases were shaded according to the level of impact that Big Data has. Table 2 indicates the colours used to shade the ADM phases.

Table 2: TOGAF ADM ranges and colours

Percentage Range	Colour
41% - 50%	
51% - 60%	
61% - 70%	
71% - 80%	
81% - 90%	
91% - 100%	
Not included in the scope	

As seen in the adapted TOGAF ADM model, the Requirements Management bubble was shaded in red indicating that it will be highly impacted by the 4 V's of Big Data. Requirements Management plays a significant role in each phase of the TOGAF ADM. The ADM is continuously driven by a requirements management process. This is a dynamic process that will be influenced by Big Data. The requirements for a data-driven enterprise differs from the requirements of an organization that is not data-driven. There is Big Data software requirements, requirements for real-time analytics, hardware requirements, etc. that will differ significantly in an data-driven organization

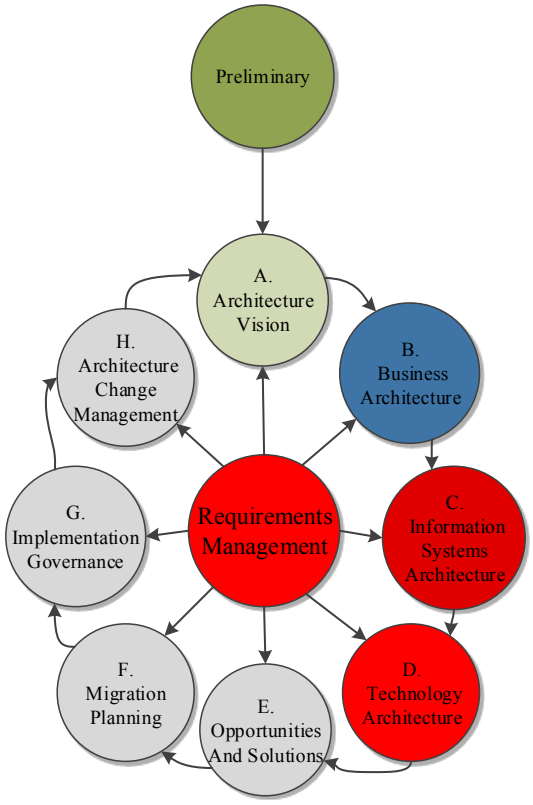

Figure 2: Adapted TOGAF ADM model

VI. CONCLUSION

In conclusion, the main problem identified is that there exists little discussion within literature with regards to the integration of data, Big Data and data science into the TOGAF ADM, specifically when considering data and big data as disruptors and the aim to be a data-driven enterprise. This in return could have a significant influence on the effectiveness as well as the competitive advantage of organizations.

The results from the systematic literature review was represented in table format to illustrate the intensity of the impact of data science on each of the first five phases of the TOGAF ADM. From that summarization table an adapted TOGAF ADM model illustrating that impact was produced as can be seen in Figure 2.

The phases of the TOGAF ADM will be impacted to a large extent if data science is taken into account. The objectives, goals, requirements, architecture, infrastructure, technology, etc. of an organization will be significantly swayed by data as shown in the Summary table (Table 1) and the above mentioned TOGAF ADM model (Figure 2).

This study has shown the importance of incorprating data science holistically into an organization and how organizations might reap significant rewards from it. Big Data and data science can be systematically incorporated using enterprise architecture, specifically the TOGAF ADM, in order to promote a more effective, efficient, and market leading organization

VII. REFERENCES

[1] J. Gantz and D. Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big Data , Bigger Digital Shadows, and Biggest Growth in the Far East," EMC Corporation. http://www.emc.com/leadership/digital-universe/2014iview/index.htm, 2012.

[2] CSC, "BIG DATA UNIVERSE BEGINNING TO EXPLODE," 2012. [Online]. Available: http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode. [Accessed: 05-Apr-2016].

[3] J. Hendler, "Big data meets computer science," *J. Comput. Sci. Coll.*, vol. 29, pp. 5–6, 2014.

[4] A. Abbasi, S. Sarker, and R. Chiang, "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," *J. Assoc. Inf. Syst.*, vol. 17, no. 2, pp. i–xxxii, 2016.

[5] S. Sagiroglu and D. Sinanc, "Big data: A review," *Int. Conf. Collab. Technol. Syst.*, pp. 42–47, 2013.

[6] N. Waller, "Get Familiar With Big Data Now—or Face 'Permanent Pink Slip,'" *Wall Street Journal*, 2014. [Online]. Available: http://online.wsj.com/news/articles/SB10001424052702304819004579489541746990638.

[7] M. A. Waller and S. E. Fawcett, "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management," *J. Bus. Logist.*, vol. 34, no. 2, pp. 77–84, 2013.

[8] The Open Group, "TOGAF." The Open Group: http://www.opengroup.org/togaf/.

[9] RealIRM, "Worldwide TOGAF Adoption Accelerates," 2010. [Online]. Available: https://www.realirm.com/articles/worldwide-togaf-adoption-accelerates. [Accessed: 01-Apr-2016].

[10] Open Group, "TOGAF ADM," *TOGAF Documentation Chapter 5*, 2011. [Online]. Available: http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap05.html. [Accessed: 01-Apr-2016].

[11] J. Zachman, "Zachman Enterprise Architecture Framework." Zachman International, p. 2011, 2011.

[12] M. de Vries, A. Gerber, and A. Van der Merwe, "The Nature of the Enterprise Engineering Discipline," *Lect. Notes Bus. Inf. Process. - Proc. EEWC 2014 - 4th Enterp. Eng. Work. Conf.*, vol. 174, pp. 1–15, 2014.

[13] H. A. E. Proper and D. D. Greefhorst, "The Roles of Principles in Enterprise Architecture," pp. 57–70, 2010.

[14] R. Villarreal, *A Systemic Perspective to Managing Complexity with Enterprise Architecture*. IGI Global, 2013.

[15] M. E. Fayad and D. S. Hamu, "Enterprise frameworks: guidelines for selection," *ACM Comput. Surv.*, vol. 32, no. 1es, p. 4–es, Mar. 2000.

[16] J. A. Zachman, "Enterprise Architecture: The Issue of the Century," *Database Program. Des. Mag.*, 1997.

[17] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. a. S. Netto, and R. Buyya, "Big Data computing and clouds: Trends and future directions," *J. Parallel Distrib. Comput.*, 2014.

[18] J. Manyika, M. Chui, J. B. R. Dobbs, P. Bisson, and A. Marrs, "McKinsey Report: Disruptive technologies: Advances that will transform life, business, and the global economy," 2013.

[19] M. Lycett, "'Datafication': making sense of (big) data in a complex world," *Eur. J. Inf. Syst.*, vol. 22, no. 4, pp. 381–386, 2013.

[20] J. Ganz and D. Reinsel, "Big Data , Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC IView*, pp. 1–16, 2012.

[21] F. J. Smith, "Data science as an academic discipline," *Data Sci. J.*, vol. 5, no. October, pp. 163–164, 2006.

[22] J. M. Stanton, *Introduction to Data science: V.3*. Syracuse University School of Information Studies, 2012.

[23] C. Okoli and K. Schabram, "Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research," *Work. Pap. Inf. Syst.*, vol. 10, no. 26, pp. 1–51, 2010.