THE TAXONOMICAL DISTINCTION BETWEEN THE CONCEPTS OF SMALL DATA AND BIG DATA

Wandisa Nyikana and Tiko Iyamu
Cape Peninsula University of Technology
Cape Town, Western Cape, South Africa

ABSTRACT

The differentiation between big data and small data is growing in confusion, in both academic and business domains. The confusion is endorsed by studies, by consistently arguing that the difference between big data and small data is not about size. The confusion between big data and small data is increasingly prohibitive for many organisations including students. As a result of the confusion and its associated implications, there are many costly decisions by individuals and groups in organisations. Consequently, productivity and quality of service are affected, which negatively reflect in competitiveness and sustainability. Thus, it is critically essential to address these growing challenges while the consequences are manageable. The objective of the study was to develop a taxonomy that distinguishes small data from big data, to remove the confusion, which often hinder understanding and affect the use for business enhancement. From the interpretivist approach perspective, the qualitative methods were employed in the study.

KEYWORDS

Big Data, Confusion between Small and Big Data, Taxonomic, Small Data

1. INTRODUCTION

Many organisations are increasingly realising the importance and value of data. The realisation triggers organisations, to gain better understanding and exploit the usefulness of their data from different angles, for effective business decisions (Ugur & Turan, 2020). The academic domain, too, are increasingly conducting studies, to gain experimental and empirical studies, to contribute to the developmental and usefulness of data (Cockcroft & Russell, 2018). In doing so, both business and academic domains characterise data into two main categories, small data and big data (Minami & Ohura, 2021; Gelhaar, Groß & Otto, 2021; Rengarajan et al., 2022). The categorisation has infused confusion for many people in both business and academics (Uğur & Turan, 2020).

Small data often refers to data or normal data. Small data is often viewed and explained as a concept that uses tiny clues and specific attributes to uncover huge trends (Rengarajan et al., 2022). Kitchin and Lauriault (2015) argue that small data is characterized by limited volume, non-continuous collection, and narrow variety. From scientific angle, Ferguson et al. (2014) explain how the small data is a collective representation of entities for various purposes. Without contradiction, small data has been used for many years by businesses, to produce meaningful insights (Kitchin & Lauriault, 2015) and to make operational decisions (Cekerevac et al., 2016).

Big data is defined by its characteristics known as the 4 V’s volume, velocity, veracity, and variety (Sun et al., 2018; Osman, 2019). According to Barham (2017), Volume refers to size, which entails the scale of data. Velocity is the speed at which data travels, including how the data or set of data is streamed and flows in exchanges (Iyamu, 2018). Veracity is the complexity and uncertainty of data (Lam et al., 2017). Variety refers to the different forms of data (Barham, 2017). According to Bariki et al. (2017), value is another characteristic that defines big data, which depends on the importance that an organisation associated with it. Small data on the other hand is the sample data retrieved by using sampling methods to understand certain problems (Cheng et al., 2018). It is characterised by its limited volume and narrow variety (Kitchin & Lauriault, 2015).

There are obvious and unclear similarities and differentiation between small data and big data. The differentiation can be clarified and put into perspectives by an understanding of the taxonomies of the concepts, which include their nomenclature. Nomenclature is the systematic way that we use to name things (Hugenholtz et al., 2021), or the rules that we use to form these names or terms (Sterner & Franz 2017). Its purpose is to provide unambiguous clear meaning of names so that there are no misunderstandings or confusion (Hawksworth, 2013). Sterner and Franz (2017) argue that nomenclature goes beyond an understanding the information that surrounds the usage of those names. Thus, standard nomenclature is required for small data and big data that can be used by both humans and machines, to gain better understanding of the concepts, idiosyncratically.

This study does not concentrate on redefining small data and big data, rather, it focuses on the confusion and distinction the concepts pose to individuals and organisations. Primarily, the confusion remains because the concepts, small data and big data are not understood, distinctively. The confusion can be attributed to lack of clarification of the taxonomies including the nomenclature of the concepts. This problem does not get easier because many studies either concentrate on big data or small data. Thus, it is hard to find studies that focus on both concepts, to increase their distinctiveness towards usefulness by organisations and stakeholders. Therefore, this study focuses on defining and establishing the taxonomies of small data and big data, for organisational purposes. This will help to provide clarity of the two concepts, eliminate the confusion and increase their usefulness.

The objective of the study was to develop a taxonomy that distinguishes small data from big data, to remove the confusion, which often hinder understanding and affect the use for business enhancement. In achieving the objective, two steps were followed in examining the phenomenon. In the first step, the nomenclatural and the differences between small data and big data were examined. Through the second step, the scope and boundaries of each concept, small data and big data were understood better. This helps to gain better understanding to, what big data is if it is not about the size. From this understanding a distinction is established.

2. PROBLEMATISATION OF THE CONCEPTS

In many organisations, the term and concept of big data remain a bass word. This is attributed to the fact that many employees or stakeholders of organisations do not seem to observe or believe that there is difference between small (normal) data and big data. In some organisations, small data is often mistaken for big data, vice versa. Consequently, this type of confusion has negative effect and influence on data structuring, management and planning for business enhancement. For example, despite the similarities, tools for big data analysis are purchased for small data purposes. In such an instance, two prohibitive things happen: (1) the cost of purchasing the tools for analysis or analytics and scarce skill required; and (2) inappropriate tool is employed, which yields undesirable results.

The small data contains some of the big data characteristics (Kitchin & Lauriault, 2015). Hence Cheng (2018) claims that big data comes from small data but did not draw boundary or distinction between the two concepts. Also, the data analytics tools used to analyse big data can also be applied to small data to extract information and gain useful insight. The overlapping of the two concepts brings confusion hence it is important to understand the nomenclature for big data and small data.

Yet, the characteristics including the nomenclature of both small data and big data are the same (Katal, Wazid & Gouda, 2013; Faraway & Augustin, 2018). Faraway and Augustin (2018) explain how the confusion makes it difficult for the data analyst/scientists to be skilled and be confident that they understand both. For examples, some organisations are challenged with pricing the services of their data because they cannot differentiate big from small; and organisations duplicate analytics tools because of lack of clarity.

3. LITERATURE REVIEW

This section presents the review of literature conducted. It focuses on the core aspects of the study, which are the small data, big data, including the differentiation between the small and big data, and the concept of taxonomy.

3.1 Small Data in Organisations

Small data is defined by Kitchin and Lauriault (2015) as sample data that is focusing at answering specific questions. It consists of structured data sets. Ahmed et al. (2017) suggest that small data is characterised by low volumes, quantified velocities, and structured varieties. Because of its manageable volumes, small data can be understood without the use of analytics (Dhaliwal & Shojania, 2018). However, low or size can be subjective if there is no universal definition or measurable agreement. Such subjectivism allows an enterprise to decide on volume (big or small) in isolation. The emergence of big data invoke contrast in the category and boundary including differentiation between the two concepts (Faraway & Augustin, 2018). Thus, it is essential to understand the characteristics and usefulness of the concepts in organisations towards to enhancing activities and improving competitiveness.

Small data focuses on discovering and understanding what causes things to happen rather than the prediction (Faraway & Augustin, 2018). Hence, it is used to determine current situations and conditions. Academic institutions specifically the researchers use small data for the intervention of the research studies. Many organisations use it to produce meaningful results and solutions (Vargas, 2018) and to discover new useful insights (Dhaliwal & Shojania, 2018). According to Cekerevac et al. (2016), organisations use it because it is properly developed and has been used for a long time. Also, it enables organisations to make key business decisions (Necsulescu, 2017). This could be attributed to the fact that small data is granular and insightful.

Small data faces challenges with machine learning algorithms from analysis perspective because small data are overfitting (Li et al., 2020; Kong et al., 2020). The machine learning algorithms do not provide robustness when applied to smaller data sets and this leads to poor performance, expensive and complex process (Kennedy et al., 2017; Vecchi et al., 2022). Also, other methods available for analysis of small data have limited effectiveness and require skilled personnel (Kong et al, 2020). On the other hand, Kennedy et al. (2017) claim that since small data is using sample data, it does not have the ability to fully represent the large data sets. Furthermore, small data focuses on answering specific questions or queries. Hence, it is difficult to apply its findings to large groups of events and activities (Ravi, 2021).

3.2 Big Data in Organisations

Big data comes from various sources with several types of data formats and structures. It is collected using different devices (Iyamu, 2020). Big data contains large, structured, semi structured, and unstructured data sets (Oussous et al., 2018). The concept is concerned with capturing, storing, analysing, and evaluating the data that is created by human beings and devices using computer technologies (Herschel & Miori, 2017).

Big data has become a crucial and useful resource to the organisations. Cockcroft and Russell (2018) highlight big data as an asset in many organisations. It is recognised in many sectors and by different professionals such as scientists and healthcare practitioners (Iyamu, 2020). Some of the organisations use it to address their processes and strategies (Barham,2017). While others use it for sustainability, efficiency, and competitiveness (Iyamu, 2018). Also, big data helps the organisations to improve decision making, to achieve their goals (Sivarajah, 2017). Moreover, it assists the organisations to understand their operations Ahmed et al. (2017) and to cut costs (Grable & Lyons, 2018). Cekerevac et al. (2016) adds that the organisations use big data to gain new insights and for prediction. Financial institutions use the big data to detect fraud (Cockcroft & Russell, 2018). While pharmaceutical companies use it to trace defects on new products (Barham, 2017).

Big data presents some challenges to organisations regardless of its usefulness. One of those challenges is the complexity with integration of the data (Barham, 2017). This is due to different data structures that big data has and the high speed in which it flows (Barham, 2017; Samsudeen & Haleem, 2020). Also, some of the organisations still have data in legacy databases and this makes it difficult to gain value from big data (Nunu,2019). According to Mgudlwa and Iyamu (2018), processing data is complex because of the large size of the data. Moreover, it is complicated to process the big data using traditional data processing applications. Nyikana and Iyamu (2022) highlight other challenges as storage, skills, searching, security, and privacy violation. On the other hand, the infrastructure for big data is inadequate and expensive, according to Sivarajah et al. (2017). Furthermore, the synchronisation of large data sets is another challenge.

3.3 Small Data and Big Data

The confusion in differentiation between small data and big data is growing and affects logic and value associated to them (Kitchin & Lauriault, 2015). Sacristán and Dilla (2015) suggest that organisations struggle in achieving the potentiality of big data because the small data are not differentiated from each other by the users. Letouzé, Areias and Jackson (2016) the dichotomy between small data versus big data does not capture the complexity of their structure and ecosystems. According to Aversa, Doherty and Hernandez (2018), currently, there seems to be no consensus on the determinants of the small data and big data. Kitchin and Lauriault (2015) posit that the term ‘big’ is misleading as big data are characterized by much more than volume, and ‘small’ data can be large, such as national censuses.

Small data and big data are usually distinguished from each other using several factors, which include scope and volume. The capability, requirements and support mechanism for small data is different from big data (Davenport, Barth & Bean, 2012). The differences draw its inferences from factors such as accessibility, conciseness, and workability. Furht and Villanustre (2016) argue that there is distinction between the small data and big data but did not detail the differences. Wang (2017) highlights heterogeneity as one of the differences between the concepts of small and big data.

3.4 The Concept of Taxonomy

It is important to categorise and classify the concept of big data and small data, which can be done through taxonomy. Rizk et al. (2018) define taxonomy as the process of classification used in scientific fields. Gelhaar et al. (2021) explain that taxonomies provide a structure and organised knowledge that can be used by the researchers to understand and analyse complex areas. Hence, developing the taxonomy of big data and small data would benefit both academics and the business domains, in gaining better insights and understanding of the existing knowledge about the concepts. Furthermore, taxonomy helps to develop rigour theory.

Taxonomies are used in the literature of information systems (IS) to analyse and classify complex phenomenon (Azkan et al., 2020). Also, to understand relationships among concepts (Rizk et al., 2018). According to Nickerson et al. (2013), there is a method that has been developed for IS researchers to use for taxonomy development to classify artifacts. Maslin (2002) explains that without taxonomy in biology, it is not easy to communicate and exchange information about organisms. Also, when the taxonomy is poorly defined, all the information linked to those defined names will be incorrect.

Taxonomy is widely used in different fields. Bloom’s taxonomy is a well-known taxonomy used in the academic domain for the classification of educational learning objectives (Anindítya et al., 2019). In chemistry, periodic table is another example where taxonomy has been used to understand the elements (Oberlander et al., 2019). While in the field of IS, taxonomy has been used for the classification of digital technologies such as the internet of things, cloud computing and social media (Berger et al., 2018; Szopinski et al., 2019). Health care uses taxonomies to classify diseases and medication to improve diagnosis (Haendel et al., 2018; Seyhan & Carini, 2019). Furthermore, in health research, taxonomy is used to categorise the results of clinical trials, to improve knowledge discovery, which makes it easier for trials in the registries and databases (Dodd et al., 2018).

4. RESEARCH METHODOLOGY

Qualitative method was employed in this study. Primarily, this is because qualitative method seeks to understand why things are the way that they are (Al-Ababneh, 2020) and the study focuses on quality rather than quantity (Iyamu & Shaanika, 2022). The method is suitable because the study seeks to understand the distinction between big data and small data, which is based on experiences, opinions, and views. That distinction cannot be discovered by using the quantitative method as the method focuses more on numbers. Another reason for using qualitative method is because it is exploratory by nature (Sovacool et al., 2018). Hence, it was used to explore the characteristics of big data and small to eliminate the confusion between the two concepts. Thus, from the perspective of qualitative, a design is selected.

Document analysis was employed in the data collection, primarily because of wide-coverage and historical purposes. According to Lakay and Iyamu (2022), documentation focuses on collecting the existing data that is stable and may sometimes not be noticeable. Furthermore, it helps to provide broad knowledge and extensive coverage of the phenomenon being studied. Iyamu, Nehemia-Maletzky and Shaanika (2016) argue that documentation helps to provide historical background of over a period. Hence, it has been adopted for the study to gain extensive and historical knowledge about small data and big data.

Criteria was set, consisting of two factors, source and period, to guide the collection of data. First, the use of academic databases, to ensure credibility and reliability of the data. Second, a period of ten years, to ensure extensive coverage of the meanings and attributes that have been associated with the concepts, historically. The data was collected from academic databases which include Google Scholar, AIS, and Ebscohost. The coverage was on articles published between 2012 and 2022. As shown in Table 1, a total of 41 articles were collected, of which 17 were for small data and 24 for big data. The Table depicts the types of materials that were collected.

Table 1. Data collection

	Journal	Conference Proceedings	Book chapter	Book	Others (e.g., white papers)	Total
Small data	8	4		1	2	15
Big data	16	5		1	1	23
Total	24	9		2	3	38

5. ANALYSING THE QUALITATIVE DATA

We are aware that conducting analysis of data in a qualitative study can be cumbersome in that there is no specific guidelines or method, as revealed and discussed in literature (Dufour & Richard, 2019; Lester, Cho & Lochmiller, 2020). Thus, we carefully and methodologically employ interpretivist approach, from whose perspective the hermeneutics approach is applied to analyse the data in this study. It was methodological in that the analysis involves a process of describing, classification and interpretation of the data, to provide relevance, and useful meaningful information (Taherdoost, 2022; Cassell & Bishop, 2019).

Hermeneutics approach is concerned with understanding and interpretation of data (Lakay & Iyamu 2022). According to Nigar (2020), the hermeneutics approach focuses on digging deep into text, to find new knowledge. Furthermore, it allows the researcher’s own understanding and interpretation of the text. According to Nyikana and Iyamu (2022), the use of the hermeneutics circles helps to gain deeper understanding of the meanings that are associated with things, through repeated reading of the texts. The circles mean continuous interrogation of the text, by going forward and back until a satisfactory point where the researchers feels that a better understanding is gained.

Based on the focus of the hermeneutics, the approach is most appropriate for this study, primarily for two reasons. First, the data is not firs-hand. Existing materials (literature) are used as data in the study, as discussed in section 6.3. This means that the researcher needs to be thorough so as to gain deeper insights of the authors of the literature’s perspectives. Second, the focus of the study, which is to determine differentiation between small data and big data is unwieldly. Therefore, it requires unfathomable details, to achieve the goal of the study. Thus, reading of the 38 (see Table 1) related materials in circles, is inevitable.

The analysis of the data is conducted in accordance with the objectives of the study, which are, to, (1) examine the nomenclature and differences between small data and big data; and (2) understand what big data is if it is not about the size.

6. DISCUSSION OF THE OUTCOME

From the analysis, there are two main outcomes, which are (1) the nomenclatural and differences between small data and big data; and (2) gain better understanding of what big data is since it is not about the size. The outcomes are presented in the remainder of this section.

6.1 The Nomenclatural and Differences between Small Data and Big Data

In understanding the nomenclature of both small data and big data, their attributes were identified, as tabulated in Table 2.

Table 2. The nomenclature

Attribute	Small data	Big data
Database	Relational Database. Data is managed and accessed using Sequel Query Language (SQL).	Non-Relational Database. Data is managed using No Sequel Query Language (No SQL)
Data Warehouse	Centralised architecture with structured datasets.	Distributed architecture with structured, semi-structured and unstructured datasets.
Data storage	Uses data marts to store the information of particular function in a single place.	Uses data lakes to store raw data from the various sources.
Data Analysis	The analysis occurs after the event. Uses statistical and business intelligence tools.	The analysis occurs in real time. Uses Big data analytics tools.

This will help to understand the scope and boundaries of each concept, small data and big data. One of the similarities is that both big data and small data contribute value to the organisations. According to Faraway and Augustin (2018), clearly, big data and small data are generated using the same sources, which include technological, business, and societal factors. Small or big data drives innovation and productivity of the businesses including decision making (Hassani & Silva, 2015). According to Jin et al. (2015), big data can enhance the competitive advantage of organisations, economic growth of countries, and help to predict the future of enterprises. Doesn’t small data provide the same capability? The differences seem to hide within each other, small data and big data.

6.2 Understand what Big Data is if it is not about the Size

Table 3 provides a distinction between small data and big data, which explains the trajectory of the concepts. The distinction shows that the differences between the concepts is beyond size.

Table 3. The characteristics

Characteristic	Small data	Big data
Volume	The data is in the range of up to hundreds of gigabytes.	The data is more is in terabytes or more.
Velocity	The data is regulated, and constantly, it flows. The aggregation of the data is slow.	The data arrives at unprecedented high speeds, and large volumes of data aggregation in a short time.
Variety	It has narrow variety of data sets which include records, files and tabular data.	It consists of numerous data set, which include tabular data, text, audio, images, video, logs, JSON.
Veracity	The data comes from reliable sources.	The data comes from several sources with complex, less reliable, bias and noise data sets.
Value	The data is used to produce insights, drive innovation, productivity and improve decision making.	The data is used to drive innovation and productivity, enhance economic growth, improve decision making and gain competitive advantage.
Structure	The data is structured, often in tabular format with fixed schema (relational).	The data is structured, semi structured, and unstructured data sets with dynamic schema (non- relational)
Discover trends/insights	The data uses tiny clues and specific attributes to discover meaningful insights. The data is current. It focuses on what causes things to happen.	The data is used to discover insights. It focuses on what causes things to happen and predict what could happen in future.
Drives decision making	The data is granular. It answers specific questions.	The data increases organisational outputs and profits.
Optimisation	The data can be optimized manually (human effort)	The data requires machine learning techniques for optimization of the data.

Based on the aim of the study which is to develop a taxonomy that distinguishes small data from big data, an in-depth investigation was conducted. The investigation focused on removing the confusion that exist between the two concepts. Thus, the two concepts, small data and big data were investigated in two phases.

First, the concepts were investigated separately. Second, the concepts were mapped against each other. This approach helps to detect the similarities and differences, towards removing the confusion between small data and big data using its characteristic, as shown in Table 3. In addition, the approach will help to gain an understanding of how factors transform to form the taxonomies of each concept.

7. CONCLUSION

The study develops a taxonomy that distinguishes small data from big data, to remove the confusion that currently exists between the two concepts. This has immense significance for business, in conducting transaction that are dependent on data and assessing value. The study is significant in IT specialists, which include managers and data architects, as they strive to support and enable organisation's aim and objectives. Through better understanding of the distinction between the concepts, data architects can design a less complex architecture from both business and technology perspectives. A fewer complex data architecture is intended to increase competitiveness and sustainability, for an organisation.

The study provides two distinctives entities, nomenclature and characteristics between small data and big data. From academics' viewpoints, each of these entities is a foundation for further development. From this perspective, the study contributes to the body of knowledge, which researchers and students, particularly, postgraduates can access for better understanding and clarifications concerning small data and big data.

The study provides clarity of an area that has been most confusing and conflict through its categorisation of the attributes and characteristics of small data and big data. This enables individuals such as data scientists and manager, data architects and organisations at large to have better understanding of the dimensions and myriad in carrying out activities such as analysis and computing of small data or big data. This can be used to define the value and contributions of either the small data or big data in an organisation. From academic viewpoint, the study can be used as baseline for developing a framework for data attribution platform. The platform will focus on business related services and value-creating mechanisms, to increase effective and efficient use of data in an organisation.

Although the study provides useful clarification for the confusion that exist between small data and big data, the work can be extended. For further studies, it will be useful and relevant to both academic and business domain, if a model is designed for the evaluation of the concepts.

REFERENCES

Ahmed, V. et al, 2017. The future of big data in facilities management: opportunities and challenges. *Facilities*, Vol. 35, No. 13, pp 725-745.

Al-Ababneh, M.M., 2020. Linking ontology, epistemology and research methodology. *Science & Philosophy*, Vol. 8, No.1, pp 75-91.

Aninditya, A. et al, 2019. Text mining approach using TF-IDF and naive Bayes for classification of exam questions based on cognitive level of bloom's taxonomy. In *International Conference on Internet of Things and Intelligence System (IoTaIS)*. Bali, Indonesia, pp. 112-117.

Aversa, J. et al, 2018. Big data analytics: "The new boundaries of retail location decision making". *Papers in Applied Geography*, Vol. 4, No. 4, pp 390-408.

Azkan, C. et al, 2020. A Taxonomy for Data-Driven Services in Manufacturing Industries. In *Twenty-Fourth Pacific Asia Conference on Information Systems.* Dubai, UAE, pp.1-14.

Barham, H., 2017. Achieving competitive advantage through big data: A literature review. In *Portland international conference on management of engineering and technology (PICMET)*. Portland, USA, pp.1-7.

Bariki, L. et al, 2017. Big data analytics: Challenges, research issues, tools and application survey. *International Journal of Engineering Sciences & Research Technology*, Vol. 6, No. 1, pp 548-554.

Berger, S. et al, 2018. The Nature of Digital Technologies-Development of a Multi-Layer Taxonomy. In *Twenty-Sixth European Conference on Information Systems*. Portsmouth, UK, pp.1-18.

Cassell, C. and Bishop, V., 2019. Qualitative data analysis: Exploring themes, metaphors and stories. *European Management Review*, Vol. 16, No. 1, pp 195-207.

Cekerevac, Z. et al, 2016. Big vs small data in micro and small companies. *Communications-Scientific letters of the University of Zilina*, Vol.18, No. 3, pp 34-40.

Cheng, J. et al, 2018. Thoughts on the problem of small data. In *IOP Conference Series: Materials Science and Engineering*. Nanjing, China, pp. 012029.

Cockcroft, S. and Russell, M., 2018. Big data opportunities for accounting and finance practice and research. *Australian Accounting Review*, Vol. 28, No. 3, pp 323-333.

Davenport, T. et al, 2012. How 'big data' is different. *MIT Sloan Management Review*, Vol. 54, No. 1, pp 21-25.

Dhaliwal, G. and Shojania, K.G., 2018. The data of diagnostic error: big, large and small. *BMJ Quality & Safety*, Vol. 27, No. 7, pp 499-501.

Dodd, S. et al, 2018. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *Journal of clinical epidemiology*, Vol. 96, No. 2018, pp 84-92.

Dufour, I.F. and Richard, M.C., 2019. Theorizing from secondary qualitative data: A comparison of two data analysis methods. *Cogent Education*, Vol. 6, No: 1-15.

Faraway, J.J. & Augustin, N.H., 2018. When small data beats big data. *Statistics & Probability Letters*, Vol. 136, No. 2018, pp 142-145.

Ferguson, A.R. et al, 2014. Big data from small data: data-sharing in the'long tail'of neuroscience. *Nature neuroscience*, Vol. 17, No. 11, pp 1442-1447.

Furht, B., and Villanustre, F., 2016. *Introduction to big data. In Big data technologies and applications*. Springer, Cham.

Gelhaar, J. et al, 2021. A taxonomy for data ecosystems. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. Hawaii, USA, pp. 6113 - 6122.

Grable, J.E. and Lyons, A.C., 2018. An Introduction to Big Data. *Journal of financial service professionals*, Vol. 72, Vol. 5, pp 17-20.

Haendel, M.A. et al, 2018. Classification, ontology, and precision medicine. *New England Journal of Medicine*, Vol. 379, No. 15, pp 1452-1462.

Hassani, H. and Silva, E.S., 2015. Forecasting with big data: A review. *Annals of Data Science*, Vol. 2, No. 1, pp 5-19.

Hawksworth D. L., 2013. *Encyclopedia of Biodiversity*. Academic Press, Waltham, Vol. 2, No. 5, pp :547-559.

Herschel, R. and Miori, V.M., 2017. Ethics & big data. *Technology in Society*, Vol 49, No. 2017, pp 31-36.

Hugenholtz, P. et al, 2021. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *The ISME Journal*, Vol. 15, No. 7, pp 1879-1892.

Iyamu, T. et al, 2016. The overlapping nature of business analysis and business architecture: What we need to know. *Electronic Journal of Information Systems Evaluation*, Vol.19, No. 3, pp 169–179.

Iyamu, T., 2018. A multilevel approach to big data analysis using analytic tools and actor network theory. *South African Journal of Information Management*, Vol. 20, No. 1, pp 1-9.

Iyamu, T., 2020. A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries. *South African Journal of Information Management*, Vol. 22, No. 1, pp 1-9.

Iyamu, T. and Shaanika, I., 2022. Assessing Business Architecture Readiness in Organisations. In *Proceedings of the 24th International Conference on Enterprise Information Systems*. Online Steaming, pp. 506-514.

Jin, X. et al, 2015. Significance and challenges of big data research. *Big data research*, Vol. 2, No. 2, pp 59-64.

Katal, A. et al, 2013. Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing*. Nodia, India, pp. 404-409.

Kennedy, O. et al, 2017. Small data. In *2017 IEEE 33rd International Conference on Data Engineering*. San Diego, CA, 1475-1476.

Kothari, C.R., 2020. Research methodology methods and techniques. 2nd ed. Daryaganj: New Age International, New Delhi.

Kitchin, R. and Lauriault, T.P., 2015. Small data in the era of big data. *GeoJournal*, Vol. 80. No. 4, pp 463-475.

Koseleva, N. and Ropaite, G., 2017. Big data in building energy efficiency: understanding of big data and main challenges. *Procedia Engineering*, Vol.172, No. 2017, pp 544-549.

Kong, S. et al, 2020. Conservative Generalisation for Small Data Analytics–An Extended Lattice Machine Approach. In *2020 International Conference on Machine Learning and Cybernetics*. Adelaide, Australia, pp. 99-104.

Kumar, R. 2018. *Research methodology: A step-by-step guide for beginners*. 5th ed. Sage Publications, London.

Lam, S.K. et al, 2017. Leveraging frontline employees' small data and firm-level big data in frontline management: An absorptive capacity perspective. *Journal of Service Research*, Vol. 20. No. 1, pp 12-28.

Lakay, D. and Iyamu, T., 2022. Examining academic performance through ant towards rpa-based system in South Africa. Education and Information Technologies, Vol. 27, No. 2022, pp 9437–9454.

Alhajj. R. and Rokne.J., *Encyclopedia on Social Network Analysis and Data Mining*. Springer, New York.

Lester, J.N. et al, 2020. Learning to do qualitative data analysis: A starting point. *Human Resource Development Review*, Vol.19, No. 1, pp 94-106.

Bamberger, M. et al, 2015. *Dealing with complexity in development evaluation: A practical approach*. SAGE. pp 221–250.

Li, Z. et al, 2020. Learning with small data. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. Houston, ACM, pp 884-887.

Maslin, B.R., 2002. The role and relevance of taxonomy in the conservation and utilisation of Australian acacias. *Conservation Science Western Australia*, Vol. 4, No. 3, pp. 1-9.

Minami, T. and Ohura, Y., 2021. Small data analysis for bigger data analysis. In *2021 Workshop on Algorithm and Big Data*. Fuzhou, China, pp. 1-8.

Mgudlwa, S. and Iyamu, T., 2018. Integration of social media with healthcare big data for improved service delivery. *South African Journal of Information Management*, Vol. 20, No. 1, pp 1-8.

Mukherjee, S.P., 2019. *A guide to research methodology: An overview of research problems, tasks and methods*. CRC Press, Boca Raton.

Nigar, N., 2020. Hermeneutic phenomenological narrative enquiry: A qualitative study design. *Theory and Practice in Language Studies*, Vol. 10, No. 1, pp 10-18.

Necsulescu, N., 2017. Focusing on small data to drive big results. *Applied Marketing Analytics*, Vol. 2, No. 4, pp 296-303.

Nickerson, R.C. et al, 2013. A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, Vol. 22, No. 3, pp 336-359.

Nyikana, W. and Iyamu, T., 2022. A Guide for selecting big data analytics tools in an organisation. In *Proceedings of the 55th Hawaii International Conference on System Sciences*. Hawaii, USA, pp. 5451-5461.

Nunu, K., 2019. A big data framework to improve government service delivery in South Africa. Unpublished Thesis, Cape Peninsula University of Technology, Cape Town. http://hdl.handle.net/20.500.11838/2962 [23 October 2019].

Oberländer, A.M. et al, 2019. Taxonomy research in information systems: a systematic assessment. In *27th European Conference on Information Systems*. Stockholm, Sweden, pp. 1 – 17.

Osman, A.M.S., 2019. A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, Vol. 91, No. 2019, pp 620-633.

Oussous, A. et al, 2018. Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, Vol. 30, No. 4, pp 431-448.

Rengarajan, S. et al, 2022. Data strategies for global value chains: Hybridization of small and big data in the aftermath of COVID-19. *Journal of Business Research*, Vol. 144, No. 2022, pp 776-787.

Rizk, A. et al, 2018. Towards a taxonomy for data-driven digital services. In *Proceedings of the 51st Hawaii international conference on system sciences*. Hawaii, USA, pp. 1076-1085.

Sacristán, J. A. and Dilla, T., 2015. No big data without small data: learning health care systems begin and end with the individual patient. *Journal of evaluation in clinical practice*, Vol 21, No. 6, pp 1014-1017.

Samsudeen, S.N. and Haleem, A., 2020. Impacts and challenges of big data: a review. *International Journal of Psychosocial Rehabilitation*, Vol. 24, No. 7, pp 479-487.

Seyhan, A.A. and Carini, C., 2019. Are innovation and new technologies in precision medicine paving a new era in patients centric care?. *Journal of translational medicine*, Vol. 17, No. 1, pp 1-28.

Sivarajah, U., 2017. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, Vol. 70, No. 2017, pp 263-286.

Sterner, B. and Franz, N.M., 2017. Taxonomy for humans or computers? Cognitive pragmatics for big data. *Biological Theory*, Vol. 12, No. 2, pp 99-111.

Sovacool, B.K., 2018. Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design. *Energy Research & Social Science*, Vol. 45, No. 2018, pp 12-42.

Sun, Z., 2018. Big data with ten big characteristics. In *Proceedings of the 2nd International Conference on Big Data Research*. Weihai, China, pp 56-61.

Szopinski, D., 2019. Because Your Taxonomy is Worth IT: towards a Framework for Taxonomy Evaluation. In *27th European Conference on Information Systems*. Stockholm, Sweden, pp 1 – 20.

Taherdoost, H., 2022. Different types of data analysis; data analysis methods and techniques in research projects. *International Journal of Academic Research in Management*, Vol. 9, No. 1, pp 1-9.

Uğur, N.G., 2020. Understanding Big Data. In Big Data Analytics for Sustainable Computing, *IGI Global*, pp 1-29. DOI: 10.4018/978-1-5225-9750-6.ch001

Vargas, L. 2018. Smart media: Museums in the new data terroir. In The Routledge Handbook of Museums, Media and Communication. Drotner, k., Dziekan, v., Parry, R. & Schrøder, K.C. (ed.). 261– 73. Routledge International Handbooks. Abingdon: Routledge: 261-273.

Vecchi, E. et al, 2022. eSPA+: Scalable Entropy-Optimal Machine Learning Classification for Small Data Problems. *Neural Computation*, Vol. 34, No. 5, pp 220-1255.

Wang, L., 2017. Heterogeneous data and big data analytics. *Automatic Control and Information Sciences*, Vol. 3, No. 1, pp 8-15.