

The logical differentiation between small data and big data

**Authors:**

Wandisa Nyikana¹ 
Tiko Iyamu¹ 

Affiliations:

¹Department of Information Technology, Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa

Corresponding author:

Wandisa Nyikana,
nyikanaw@cput.ac.za

Dates:

Received: 12 Apr. 2023

Accepted: 03 Oct. 2023

Published: 24 Dec. 2023

How to cite this article:

Nyikana, W. & Iyamu, T., 2023, 'The logical differentiation between small data and big data', *South African Journal of Information Management* 25(1), a1701. <https://doi.org/10.4102/sajim.v25i1.1701>

Copyright:

© 2023. The Authors.
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Background: The distinction between small data and big data is increasingly muted and has caused challenges and confusion in many quarters.

Objective: The objective of the study is to gain a deeper understanding of the confounded confusion that exists between small data and big data. Firstly, to develop a taxonomy that distinguishes between small data and big data. Secondly, it seeks to extract the value from the concepts, which can be of fundamental importance to an organisation.

Methods: This study follows the interpretive approach and employs qualitative methods, based on which 57 related materials were gathered, covering big data and small data, and analysed.

Results: The study reveals the factors that differentiate the concepts, which are of a technical front, business logic and data processing.

Conclusion: This study addresses the challenges which are increasingly of prohibitive ramifications for both academic and business domains. By removing the confusion, the classifications of small data and big data including associated attributes will be better understood. This increases their business use towards enhancement and competitive advantage.

Contribution: The article distinguishes between small data and big data, which has been missing, from both academic and business perspectives, since the emergence of the latter. The differentiation between small data and big data provides a guide to organisations in developing strategic frameworks and operational plans.

Keywords: big data; confusion between small and big data; taxonomic; small data; data differentiation; data classification; data nomenclature; data characteristics; logic difference; data analytics.

Introduction

Many organisations are increasingly realising the importance and value of data. The realisation triggers organisations, to gain a better understanding and exploit the usefulness of their data from different angles, for effective business decisions (Ugur & Turan 2020). The academic domain, too, is increasingly conducting studies, to gain experimental and empirical studies, to contribute to the development 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 (Gelhaar, Groß & Otto 2021; Minami & Ohura 2021). The categorisation has infused confusion for many people in both business and academics (Rengarajan et al. 2022).

Small data often refers to data or normal data. Also, 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 a 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 4Vs: volume, velocity, veracity and variety (Sun, Strang & Li 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).

Read online:

Scan this QR code with your smart phone or mobile device to read online.

According to Bariki, Arvind and Hari (2017), value is another characteristic that defines big data, which depends on the importance an organisation associates with it. Small data, on the other hand, is the sample data retrieved by using sampling methods to understand certain problems (Cheng, Chen & Gong 2018). It is characterised by its limited volume and narrow variety (Kitchin & Lauriault 2015).

There are obvious and unclear similarities and differences between small data and big data. The differentiation can be clarified and put into perspective by a deeper understanding of the taxonomies of the concepts, which include their nomenclature. Nomenclature is the systematic way that is employed in naming things (Hugenholtz et al. 2021), or the rules that are used to form names or terms (Sternier & Franz 2017). Its purpose is to provide unambiguous clear meanings of names, to avoid misunderstandings or confusion. Sternier and Franz (2017) argue that nomenclature goes beyond 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 a better understanding of the concepts, idiosyncratically.

This study does not intend to redefine the concepts of small data and big data; rather, it focuses on the confusion and distinction the concepts pose to individuals and organisations (Nyikana & Iyamu 2023a). Primarily, the confusion remains because the concepts of small data and big data are not understood, distinctively. The confusion can be attributed to a 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 confusion and increase their usefulness.

Classification between small data and big data is increasingly challenging, in both academic and business domains. It is hard to find studies that explicitly clarify the confusion between big data and small data, empirically or by experiment. Also, studies have identified the confusion and consistently stating or narrowing the difference between big data and small data, not about size. The confusion between big data and small data is increasingly prohibitive for many organisations including students at all levels. As a ramification of the confusion and its associated challenges, many decisions of a costly nature are being made by individuals and groups in organisations. Consequently, some of the areas that are negatively affected include data as a service, data manipulation and quality of service, which reflect in service value, competitiveness and sustainability. Thus, it is critically essential to address these growing challenges while the consequences are manageable.

In many organisations, the term and concept of big data remain a buzzword. This is attributed to the fact that many employees or stakeholders of organisations do not seem to observe or believe that there is a difference between small (normal) data and big data. In some organisations, small data is often mistaken for big data, and *vice versa*. Consequently, this type of confusion has a 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 the scarce skill required (Mustapha 2022; Vassakis, Petrakis & Kopanakis 2018) and (2) an inappropriate tool is employed, which yields undesirable results (Kangelani & Iyamu 2020).

The small data contains some of the big data characteristics (Kitchin & Lauriault 2015). Hence, Cheng et al. (2018) claim that big data comes from small data but does not draw boundaries or distinctions between the two concepts. Also, the data analytics tools used to analyse big data can be applied to small data to extract information and gain useful insight. The overlapping of the two concepts induces more 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 and Augustin (2018) explain how the confusion makes it difficult for both data analysts and data scientists to be skilled and confident that they have a good understanding of small data and big data. For example, some organisations are challenged with pricing the services of their data because they cannot differentiate small data from big data. Also, some organisations duplicate analytics tools because there is a lack of clarity.

The objectives of the study are twofold. Firstly, it is to develop a taxonomy that distinguishes between small data from big data, to remove the confusion, which often hinders understanding of their classification for business use towards enhancement. Secondly, it is to identify the value that the logical distinction can add to an organisation. In achieving the first objective, two steps were followed in examining the phenomenon: In step 1, the nomenclatural and the differences between small data and big data were examined; and in step 2, the scope and boundaries of each concept, small data and big data were understood better. This helps to gain a better understanding of what big data is if it is not about the size. From this understanding, a distinction is established. In the second objective, heuristics is applied, to extract the intended usefulness, which can be of fundamental value to organisations.

Literature review

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

Small data in organisations

Simplistically, Kitchin and Lauriault (2015) explained that small data uniquely focuses on 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 invokes 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 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. In many academic institutions, some researchers make use of small data to assess and evaluate research outputs. Also, many organisations use small data to produce meaningful results and solutions (Vargas 2018) and to discover new useful insights (Dhaliwal & Shojania 2018). According to Cekerevac et al. (2016), organisations employ small data because it is most appropriate for developing initiatives. 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.

From an analysis viewpoint, the use of small data with machine learning algorithms faces challenges because can be overfitting (Kong, Wang & Wang 2020; Li, Yao & Ma 2020). The machine learning algorithms do not provide robustness when applied to smaller data sets and this leads to poor performance, and expensive and complex processes (Kennedy et al. 2017; Vecchi et al. 2022). Also, other methods available for the 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 uses sample data, it cannot fully represent 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).

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 for 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 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 organisations to improve decision-making, to achieve their goals (Sivarajah et al. 2017). Moreover, it assists organisations to understand their operations (Ahmed et al. 2017) and to cut costs (Grable & Lyons 2018). Cekerevac et al. (2016) add that organisations use big data to gain new insights and for prediction. Financial institutions use big data to detect fraud (Cockcroft & Russell 2018), while pharmaceutical companies use it to trace defects in new products (Barham 2017).

Big data presents some challenges to organisations regardless of its usefulness. One of those challenges is the complexity of the integration of the data (Barham 2017). This is due to the different data structures that big data has and the high speed at 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. According to Mgudlwa and Iyamu (2018), processing data is complex because of the large size of the data. Moreover, it is complicated to process big data using traditional data processing applications. Nyikana and Iyamu (2022) highlight other challenges such as storage, skills, searching, security and privacy violations. 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 (Nyikana & Iyamu 2023b).

Small data and big data differentiation

The confusion in differentiation between small data and big data is growing and it affects the logic and value associated with them (Kitchin & Lauriault 2015). Sacristán and Dilla (2015) suggest that organisations struggle to achieve the potential of big data and small data because many users find it difficult to differentiate between the attributes of the concepts. According to Letouze, Areias and Jackson (2015), the dichotomy between small data versus big data does not capture the complexity of their structures and ecosystems. Currently, there seems to be no consensus on the determinants of small data and big data (Aversa, Doherty & Hernandez 2018; Nyikana & Iyamu 2023a). In an attempt to gain an understanding of the challenges, 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 mechanisms for small data are different from big data (Davenport, Barth & Bean 2012). The differences draw inferences from factors such as accessibility, conciseness and workability. Furht and

Villanustre (2016) argue that there is a distinction between small data and big data but do not detail the differences. Wang (2017) highlights heterogeneity as one of the differences between the concepts of small and big data.

The concept of taxonomy

It is important to categorise and classify the concepts of big data and small data, which can be done through taxonomy. Rizk, Bergvall-Kåreborn and Elragal (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 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 theory.

Taxonomies are used in the literature about information systems (IS) to analyse and classify complex phenomena (Azkan et al. 2020). Also, to understand relationships among concepts (Rizk et al. 2018). According to Nickerson, Varshney and Muntermann (2013), there is a method that has been developed for IS researchers to use for taxonomy development to classify artefacts. 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 (Gkinko & Elbanna 2023; Prudencio, Maximo & Colombini 2023).

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 (Aninditya, Hasibuan & Sutoyo 2019). In chemistry, the periodic table is another example where taxonomy has been used to understand the elements (Oberländer, Lösner & Rau 2019). In the field of IS, taxonomy has been used for the classification of digital technologies such as the Internet of Things (IoT), cloud computing and social media (Berger, Denner & Roeglinger 2018; Szopinski, Schoormann & Kundisch 2019). Healthcare uses taxonomies to classify diseases and medication to improve diagnosis (Haendel, Chute & Robinson 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).

Research methods

A qualitative method was employed in this study. Primarily, this is because the 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 the qualitative method is that it is exploratory by nature (Sovacool, Axsen & Sorrell 2018). Hence, it was used to explore the characteristics of big data and small data, to eliminate the confusion between the two concepts.

Document analysis was employed in the data collection, primarily because of wide coverage and historical purposes. According to Lakay and Iyamu (2022), the 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 the document analysis approach helps to provide balance and historical background over a period. Thus, the approach was employed to gain extensive and historical knowledge about big data and small data.

Criteria were set, consisting of two factors; source and period, to guide the collection of data. Firstly, the use of academic databases, to ensure credibility and reliability of the data. Secondly, a period of 10 years, to ensure extensive coverage of the meanings and attributes that have been associated with the concepts, historically. The data were collected from academic databases that include Google Scholar, AIS and Ebscohost. Over 450 articles were collected, requiring formulation and use of criteria. The criteria are as follows: (1) an article covering big data, small data, definition, and contrast between the concepts; and (2) an article published between 2012 and 2022. As shown in Table 1, the most appropriate articles were narrowed to a total of 57, of which 21 and 36 were for small data and big data, respectively.

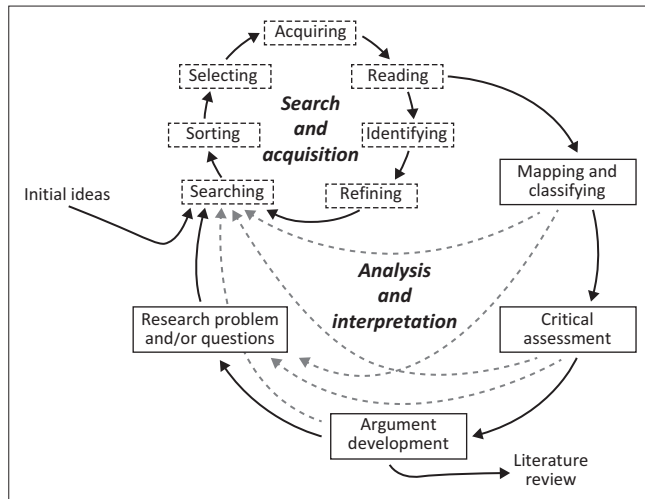
Analysing the qualitative data

We are aware that analysing data in a qualitative study can be cumbersome in that there are no specific guidelines or methods, as revealed and discussed in the literature (Dufour & Richard 2019; Lester, Cho & Lochmiller 2020). Thus, we carefully and methodologically employ the interpretive 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 relevant, useful and meaningful information (Cassell & Bishop 2019; Taherdoost 2022). The authors adopted the framework by Boell and Cezec-Kecmanovic (2014), shown in Figure 1, for the analysis.

The 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

TABLE 1: Source of data.

Concept	Journal	Conference proceedings	Book	Others (e.g. white papers)	Total
Small data	10	8	1	2	21
Big data	23	11	1	1	36
Total	33	19	2	3	57



Source: Boell, S.K. & Cezec-Kecmanovic, D., 2014, 'A hermeneutic approach for conducting literature reviews and literature searches', *Communications of the Association for Information Systems* 34(12), 257–286. <https://doi.org/10.17705/1CAIS.03412>

FIGURE 1: Hermeneutic Framework.

the researcher's understanding and interpretation of the text. According to Nyikana and Iyamu (2022), the use of hermeneutics circles helps to gain a 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 feel 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. Firstly, the data are not first-hand. Existing materials (literature) are used as data in the study, as discussed in 'Processing mechanism' section. This means that the researcher needs to be thorough, to gain deeper insights into the authors of the literature's perspectives. Secondly, the focus of the study, which is to determine the differentiation between small data and big data is unwieldy. 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 data analysis is conducted based on 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.

Finding the distinction between the concepts

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

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.

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 organisations. According to Faraway and Augustin (2018), 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 businesses including decision-making (Hassani & Silva 2015). According to Jin et al. (2015), big data can enhance the competitive advantage of organisations, and 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.

What is big data if it is not about 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 are beyond size.

TABLE 2: The nomenclature.

Attribute	Small data	Big data
Database	Relational database. Data are 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 a particular function in a single place.	Uses data lakes to store raw data from 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.

TABLE 3: The characteristics.

Characteristic	Small data	Big data
Volume	The data are in the range of up to hundreds of gigabytes.	The data are more in terabytes or more.
Velocity	The data are regulated and constantly flows. The aggregation of the data is slow.	The data arrive at unprecedented high speeds and large volumes of data aggregation in a short time.
Variety	It has a narrow variety of data sets, which include records, files and tabular data.	It consists of numerous data sets, which include tabular data, text, audio, images, video, logs and JSON.
Veracity	The data come from reliable sources.	The data come from several sources with complex, less reliable, biased and noisy data sets.
Value	The data are used to produce insights, drive innovation, increase productivity and improve decision-making.	The data are used to drive innovation and productivity, enhance economic growth, improve decision making and gain competitive advantage.
Structure	The data are structured, often in tabular format with fixed schema (relational).	The data are structured, semi-structured and unstructured data sets with dynamic schema (non-relational)
Discover trends and insights	The data uses tiny clues and specific attributes to discover meaningful insights. The data are current. It focuses on what causes things to happen.	The data are used to discover insights. It focuses on what causes things to happen and predicts what could happen in future.
Drives decision making	The data are granular. It answers specific questions.	The data increase organisational outputs and profits.
Optimisation	The data can be optimized manually (human effort).	The data require machine learning techniques for optimisation 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 exists between the two concepts. Thus, the two concepts – small data and big data – were investigated in two phases. Firstly, the concepts were investigated separately. Secondly, 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 characteristics, 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.

Ethical considerations

Ethical clearance was obtained from the Faculty of Informatics and Design (FID) Ethics Committee, Cape Peninsula University of Technology. It does not have project research number, but student number: 203168283.

Discussion of the differentiation

The study examined the small data and big data to gain a better understanding and taxonomic distinction between the two entities. The following discussion is based on the summation of the analysis of Table 2 and Table 3, which focused on the nomenclature and characteristics of small data and big data. From the tables, three factors of fundamental values are subjectively extracted, which gives credence to the distinction between the two concepts. The distinctive values are the technical front, business logic and data processing perspectives.

Technical front

From the technical front, small data and big data can be distinguished from three significant perspectives: architecture, storage and database. The architectural design provides 'fit' for both small data and big data, to enhance usefulness for organisational purposes. Although they are analogous, the architecture for small data and big data is not the same. Small data has a centralised architecture with structured data sets. According to Ahmed et al. (2017), the centralised architecture is sufficient for the small data because the data are in a central location. Big data encompasses distributed architecture with structured, semi-structured and unstructured data sets. The distributed architecture provides the capability to compile massive and heterogeneous data sets from various sources (Wang et al. 2020).

Furthermore, the storage of small data requires marts and stores data on a single entity or line-of-business (LOB), which can be limiting to business users (Jameel, Adil & Bahjat 2022; Najm et al. 2022). Hamoud et al. (2021) suggest that the use of data marts is limited because it is specific in terms of type and location of storage. Whereas data lakes is often used in storing big data, from various sources. Thus, big data can be stored in data lakes, in its original format and can only be

transformed when it is needed (Mathis 2017). This eliminates the pre-processing and transformation of the data before it is loaded into a data warehouse (Khine & Wang 2018). From the database point of view, the small data uses a relational database; it is managed and accessed using a sequence query language. In its divergence, big data uses non-relational databases to manage the data. Wang et al. (2020) explain that the non-relational database is suitable for the characteristics of big data, which are the huge volume, high velocity and the variety of data. In addition, it provides highly scalable and reliable data storage.

Business logic

Business logic is an expression of organisational vision and strategy. Ruiz and Gandia (2022) suggest that business logic implies a logic of service in an organisation. Thus, it is the primary determinant of, the accumulation, manipulation, and use of small data and big data. According to Gerlitz, Gerken and Hülsbeck (2023), business logic dominates sustainability and strategic processes. In return, both small data and big data add value to the business of an organisation. Some of the values are, to gain insights, drive innovation and improve decision-making towards enhancing efficiency and effectiveness (Hanafizadeh et al. 2021; Vasile & Simion 2021). Primarily, this is to fortify sustainability and increase competitive advantage. From small data, insights are discovered using current data to understand what causes things to happen in the way that they do. Ahmed et al. (2017) state that the small data focuses on answering specific questions and addressing problems. Conversely, big data helps to comprehend the current state and to predict what could happen in the future. This helps an organisation to identify deficiencies, and threats to sustainability, as well as future opportunities and growth of the business. Furthermore, Custers and Uršič (2016) argue that the most innovative and profitable businesses have based their strategy on gathering and utilising big data.

Processing mechanism

Small data flows at a slower, constant speed as compared with big data. Also, small data is granular, meaning it can be processed and analysed, manually. Also, small data is often gathered on a human scale using a personal computer, with the focus on gaining an understanding of the causation (Faraway & Augustin 2018). However, statistical tools and business intelligence tools can also be used to optimise small data, where and when necessary. In contrast, big data flows at a higher speed and often lacks constancy in its velocity. Owing to its characteristics, it is difficult, near impossible to manually analyse big data. Thus, Kitchin and Lauriault (2015) argued the need for computational algorithms. This brings about big data analytics tools and machine learning techniques to process and analyse the data. Big data analytics tools can analyse diverse and versatile big datasets and extract useful information (Nyikana & Iyamu 2022). The aforementioned analysis reveals that the taxonomic distinction between small data and big data is not only about size, but about fundamental

factors and values, such as data storage, database, data warehouse, data structure and data analysis.

Conclusion

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

The study provides two distinctive values from nomenclature and characteristics entities, between small data and big data. From an academic viewpoint, 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 on an area that has been most confusing and conflicting through its categorisation of the attributes and characteristics of small data and big data. This enables individuals such as data scientists and managers, data architects and organisations at large to have a 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 an academic viewpoint, the study can be used as a baseline for developing a framework for data attribution platforms. 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 exists 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 domains, if a classification model is designed for the evaluation of the concepts.

Acknowledgements

This article is based on the thesis of the first author's (Wandisa Nyikana) degree of Postgraduate diploma in Information Technology, Cape Peninsula University of Technology. The second author (Tiko Iyamu) was the supervisor of the project.

The authors would like to thank the Department of Information Technology, Cape Peninsula University

Technology for its support. To our colleagues in the research forum, we appreciate your support.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

W.N. and T.I. are the only two authors. W.N. problematised the topic. T.I. guided W.N. in the data collection process. Both authors conducted the analysis of the data. Both authors contributed in writing the research report.

Funding information

The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability

Access to academic journals was granted using student number. Access to documentation (peer-reviewed articles) was obtained from academic journals the university subscribes to. The authors of the materials used were clearly acknowledged through referencing.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

References

- Ahmed, V., Tezel, A., Aziz, Z. & Sibley, M., 2017, 'The future of big data in facilities management: Opportunities and challenges', *Facilities* 35(13), 725–745. <https://doi.org/10.1108/F-06-2016-0064>
- Al-Ababneh, M.M., 2020, 'Linking ontology, epistemology and research methodology', *Science & Philosophy* 8(1), 75–91.
- Aninditya, A., Hasibuan, M.A. & Sutoyo, E., 2019, 'Text mining approach using TF-IDF and naïve Bayes for classification of exam questions based on cognitive level of bloom's taxonomy', in *International Conference on Internet of Things and Intelligence System (IoT&IS)*, IEEE, Bali, November 05–07, 2019.
- Aversa, J., Doherty, S. & Hernandez, T., 2018, 'Big data analytics: The new boundaries of retail location decision making', *Papers in Applied Geography* 4(4), 390–408. <https://doi.org/10.1080/23754931.2018.1527720>
- Azkan, C., Iggena, L., Gür, I., Möller, F. & Otto, B., 2020, 'A taxonomy for data-driven services in manufacturing industries', in *Twenty-fourth Pacific Asia Conference on Information Systems, PACIS*, Dubai, June 20–24, 2020.
- Barham, H., 2017, 'Achieving competitive advantage through big data: A literature review', in *Portland international conference on management of engineering and technology (PICMET)*, IEEE, Portland, OR, July 09–13, 2017.
- Bariki, L., Arvind, T. & Hari, S.R., 2017, 'Big data analytics: Challenges, research issues, tools and application survey', *International Journal of Engineering Sciences & Research Technology* 6(11), 548–554.
- Berger, S., Denner, M.S. & Roeglinger, M., 2018, 'The nature of digital technologies-development of a multi-layer taxonomy', in *Twenty-sixth European conference on information systems*, ECIS, Portsmouth, June 23–28, 2018.
- Boell, S.K. & Cezec-Kecmanovic, D., 2014, 'A hermeneutic approach for conducting literature reviews and literature searches', *Communications of the Association for Information Systems* 34(12), 257–286. <https://doi.org/10.17705/1CAIS.03412>
- Cassell, C. & Bishop, V., 2019, 'Qualitative data analysis: Exploring themes, metaphors and stories', *European Management Review* 16(1), 195–207. <https://doi.org/10.1111/emre.12176>

- Cekerevac, Z., Dvorak, Z., Prigoda, L. & Cekerevac, P., 2016, 'Big vs small data in micro and small companies', *Communications-Scientific Letters of the University of Zilina* 18(3), 34–40. <https://doi.org/10.26552/com.C.2016.3.34-40>
- Cheng, J., Chen, W. & Gong, Y., 2018, 'Thoughts on the problem of small data', in *IOP conference series: Materials science and engineering*, IOP: 012029, Nanjing, August 17–19, 2018.
- Cockcroft, S. & Russell, M., 2018, 'Big data opportunities for accounting and finance practice and research', *Australian Accounting Review* 28(3), 323–333. <https://doi.org/10.1111/auar.12218>
- Custers, B. & Uršič, H., 2016, 'Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection', *International Data Privacy Law* 6(1), 4–15. <https://doi.org/10.1093/idpl/ipv028>
- Davenport, T.H., Barth, P. & Bean, R., 2012, 'How "big data" is different', *MIT Sloan Management Review* 54(1), 21–25.
- Dhaliwal, G. & Shojania, K.G., 2018, 'The data of diagnostic error: Big, large and small', *BMJ Quality & Safety* 27(7), 499–501. <https://doi.org/10.1136/bmjqs-2018-007917>
- Dodd, S., Clarke, M., Becker, L., Mavergames, C., Fish, R. & Williamson, P.R., 2018, 'A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery', *Journal of Clinical Epidemiology* 96, 84–92. <https://doi.org/10.1016/j.jclinepi.2017.12.020>
- Dufour, I.F. & Richard, M.C., 2019, 'Theorizing from secondary qualitative data: A comparison of two data analysis methods', *Cogent Education* 6(1), 1–15. <https://doi.org/10.1080/2331186X.2019.1690265>
- Faraway, J.J. & Augustin, N.H., 2018, 'When small data beats big data', *Statistics & Probability Letters* 136(2018), 142–145. <https://doi.org/10.1016/j.spl.2018.02.031>
- Ferguson, A.R., Nielson, J.L., Cragin, M.H., Bandrowski, A.E. & Martone, M.E., 2014, 'Big data from small data: Data-sharing in the 'long tail' of neuroscience', *Nature Neuroscience* 17(11), 1442–1447. <https://doi.org/10.1038/nn.3838>
- Furht, B. & Villanustre, F., 2016, 'Introduction to big data', in B. Furht & F. Villanustre (eds.), *Big data technologies and applications*, pp. 3–11, Springer, Cham.
- Gelhaar, J., Groß, T. & Otto, B., 2021, 'A taxonomy for data ecosystems', in *Proceedings of the 54th Hawaii international conference on system sciences*, HICSS, Hawaii, January 5–8, 2021.
- Gerlitz, A., Gerken, M. & Hülsbeck, M., 2023, 'We are a family, not a charity – How do family and business logics shape environmental sustainability strategies? A cross-sectional qualitative study', *Journal of Cleaner Production* 413, 137426. <https://doi.org/10.1016/j.jclepro.2023.137426>
- Gkinko, L. & Elbanna, A., 2023, 'The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users', *International Journal of Information Management* 69, 102568. <https://doi.org/10.1016/j.ijinfomgt.2022.102568>
- Grable, J.E. & Lyons, A.C., 2018, 'An introduction to big data', *Journal of Financial Service Professionals* 72(5), 17–20.
- Haendel, M.A., Chute, C.G. & Robinson, P.N., 2018, 'Classification, ontology, and precision medicine', *New England Journal of Medicine* 379(15), 1452–1462. <https://doi.org/10.1056/NEJMr1615014>
- Hamoud, A.K., Marwah, M.H., Alhilfi, Z. & Sabr, R.H., 2021, 'Implementing data-driven decision support system based on independent educational data mart', *International Journal of Electrical and Computer Engineering* 11(6), 5301–5314. <https://doi.org/10.11591/ijece.v11i6.pp5301-5314>
- Hanafizadeh, P., Hatami, P., Analoui, M. & Albadvi, A., 2021, 'Business model innovation driven by the Internet of Things technology, in Business service providers' business context', *Information Systems and e-Business Management* 19(4), 1175–1243. <https://doi.org/10.1007/s10257-021-00537-0>
- Hassani, H. & Silva, E.S., 2015, 'Forecasting with big data: A review', *Annals of Data Science* 2(1), 5–19. <https://doi.org/10.1007/s40745-015-0029-9>
- Herschel, R. & Miori, V.M., 2017, 'Ethics & big data', *Technology in Society* 49, 31–36. <https://doi.org/10.1016/j.techsoc.2017.03.003>
- Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D.H. & Soo, R.M., 2021, 'Prokaryotic taxonomy and nomenclature in the age of big sequence data', *The ISME Journal* 15(7), 1879–1892. <https://doi.org/10.1038/s41396-021-00941-x>
- Iyamu, T., 2018, 'A multilevel approach to big data analysis using analytic tools and actor-network theory', *South African Journal of Information Management* 20(1), 1–9. <https://doi.org/10.4102/sajim.v20i1.914>
- Iyamu, T., 2020, 'A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries', *South African Journal of Information Management* 22(1), 1–9. <https://doi.org/10.4102/sajim.v22i1.1117>
- Iyamu, T., Nehemia-Maletsky, M. & Shaanika, I., 2016, 'The overlapping nature of business analysis and business architecture: What we need to know', *Electronic Journal of Information Systems Evaluation* 19(3), 169–179.
- Iyamu, T. & 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, April 25–27, 2022.
- Jameel, K., Adil, A. & Bahjat, M., 2022, 'Analyses the performance of data warehouse architecture types', *Journal of Soft Computing and Data Mining* 3(1), 45–57.
- Jin, X., Wah, B.W., Cheng, X. & Wang, Y., 2015, 'Significance and challenges of big data research', *Big Data Research* 2(2), 59–64. <https://doi.org/10.1016/j.bdr.2015.01.006>
- Kangelani, P. & Iyamu, T., 2020, 'A model for evaluating big data analytics tools for organisation purposes', in *Responsible design, implementation and use of information and communication technology: 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society*, pp. 493–504, Springer International Publishing, Skukuza, April 06–08, 2020.
- Katal, A., Wazid, M. & Goudar, R.H., 2013, 'Big data: Issues, challenges, tools and good practices', in *2013 Sixth international conference on contemporary computing*, IEEE, Nodia, August 08–10, 2013.
- Kennedy, O., Hipp, D.R., Idreos, S., Marian, A., Nandi, A., Troncoso, C. et al., 2017, 'Small data', in *2017 IEEE 33rd international conference on data engineering*, IEEE, San Diego, CA, 19–22 April, 2017.
- Khine, P.P. & Wang, Z.S., 2018, 'Data lake: A new ideology in big data era', in K. Eguchi & T. Chen (eds.), *ITM web of conferences*, vol. 17, p. 03025, EDP Sciences.
- Kitchin, R. & Lauriault, T.P., 2015, 'Small data in the era of big data', *GeoJournal* 80(4), 463–475. <https://doi.org/10.1007/s10708-014-9601-7>
- Kong, S., Wang, H. & Wang, K., 2020, 'Conservative generalisation for small data analytics – An extended lattice machine approach', in *2020 International conference on machine learning and cybernetics*, Springer, Adelaide, December 02, 2020.
- Lakay, D. & Iyamu, T., 2022, 'Examining academic performance through ant towards rpa-based system in South Africa', *Education and Information Technologies* 27(2022), 9437–9454. <https://doi.org/10.1007/s10639-022-11007-6>
- Lam, S.K., Sleep, S., Hennig-Thurau, T., Sridhar, S. & Saboo, A.R., 2017, 'Leveraging frontline employees' small data and firm-level big data in frontline management: An absorptive capacity perspective', *Journal of Service Research* 20(1), 12–28. <https://doi.org/10.1177/1094670516679271>
- Lester, J.N., Cho, Y. & Lochmiller, C.R., 2020, 'Learning to do qualitative data analysis: A starting point', *Human Resource Development Review* 19(1), 94–106. <https://doi.org/10.1177/1534484320903890>
- Letouze, E., Areias, A. & Jackson, S., 2015, 'The valuation of complex development interventions in the age of big data', in M. Bamberger, J. Vaessen & E. Raimondo (eds.), *Dealing with complexity in development evaluation: A practical approach*, pp. 221–250, Sage.
- Li, Z., Yao, H. & Ma, F., 2020, 'Learning with small data', in *Proceedings of the 13th international conference on web search and data mining*, pp. 884–887, ACM, Houston, TX, February 03–07, 2020.
- Maslin, B.R., 2002, 'The role and relevance of taxonomy in the conservation and utilisation of Australian acacias', *Conservation Science Western Australia* 4(3), 1–9.
- Mathis, C., 2017, 'Data lakes', *Datenbank-Spektrum* 17(3), 89–293. <https://doi.org/10.1007/s13222-017-0272-7>
- Mgudlwa, S. & Iyamu, T., 2018, 'Integration of social media with healthcare big data for improved service delivery', *South African Journal of Information Management* 20(1), 1–8. <https://doi.org/10.4102/sajim.v20i1.894>
- Minami, T. & Ohura, Y., 2021, 'Small data analysis for bigger data analysis', in *2021 workshop on algorithm and big data*, ACM, Fuzhou, March 12–14, 2021.
- Mustapha, S.S., 2022, 'The UAE employees' perceptions towards factors for sustaining big and continuous impact on their organization's performance', *Sustainability* 14(22), 15271. <https://doi.org/10.3390/su142215271>
- Najm, I.A., Dahr, J.M., Hamoud, A.K., Alasady, A.S., Awadh, W.A., Kamel, M.B., et al., 2022, 'OLAP mining with educational data mart to predict students' performance', *Informatica* 46(5), 11–19. <https://doi.org/10.31449/inf.v46i5.3853>
- Neculescu, N., 2017, 'Focusing on small data to drive big results', *Applied Marketing Analytics* 2(4), 296–303.
- Nickerson, R.C., Varshney, U. & Muntermann, J., 2013, 'A method for taxonomy development and its application in information systems', *European Journal of Information Systems* 22(3), 336–359. <https://doi.org/10.1057/ejis.2012.26>
- Nigar, N., 2020, 'Hermeneutic phenomenological narrative enquiry: A qualitative study design', *Theory and Practice in Language Studies* 10(1), 10–18. <https://doi.org/10.17507/tpls.1001.02>
- Nyikana, W. & 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*, HICSS, Hawaii, January 04–07, 2022.
- Nyikana, W. & Iyamu, T., 2023a, 'The taxonomical distinction between the concepts of small data and big data', in *Proceedings of the 16th IADIS international conference information systems*, International Association for Development of the Information Society Press, Lisbon, March 11–13, 2023.
- Nyikana, W. & Iyamu, T., 2023b, 'A formulaic approach for selecting big data analytics tools for organizational purposes', in Z. Sun (ed.), *Handbook of research on driving socioeconomic development with big data*, pp. 224–242, IGI Global, Hershey.
- Oberländer, A.M., Lösser, B. & Rau, D., 2019, 'Taxonomy research in information systems: A systematic assessment', in *27th European conference on information systems*, ECIS, Stockholm, June 08–14, 2019.
- Osman, A.M.S., 2019, 'A novel big data analytics framework for smart cities', *Future Generation Computer Systems* 91(2019), 620–633. <https://doi.org/10.1016/j.future.2018.06.046>
- Oussou, A., Benjelloun, F.Z., Lahcen, A.A. & Belfkih, S., 2018, 'Big data technologies: A survey', *Journal of King Saud University-Computer and Information Sciences* 30(4), 431–448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Prudencio, R.F., Maximo, M.R. & Colombini, E.L., 2023, 'A survey on offline reinforcement learning: Taxonomy, review, and open problems', *IEEE Transactions on Neural Networks and Learning Systems*, 1–23.
- Ravi, A., 2021, 'If we didn't solve small data in the past, how can we solve Big Data today?', *arXiv preprint arXiv:2111.04442*.
- Rengarajan, S., Narayanamurthy, G., Moser, R. & Pereira, V., 2022, 'Data strategies for global value chains: Hybridization of small and big data in the aftermath of COVID-19', *Journal of Business Research* 144(2022), 776–787. <https://doi.org/10.1016/j.jbusres.2022.02.042>
- Rizk, A., Bergvall-Kärebörn, B. & Elragal, A., 2018, 'Towards a taxonomy for data-driven digital services', in *Proceedings of the 51st Hawaii international conference on system sciences*, HICSS, Hawaii, January 03–06, 2018.

- Ruiz, É. & Gandia, R., 2022, 'The key role of the event in combining business and community-based logics for managing an ecosystem: Empirical evidence from Lyon e-Sport', *European Management Journal* 41(4), 560–574. <https://doi.org/10.1016/j.emj.2022.07.005>
- Sacristán, J.A. & 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* 21(6), 1014–1017. <https://doi.org/10.1111/jep.12350>
- Samsudeen, S.N. & Haleem, A., 2020, 'Impacts and challenges of big data: A review', *International Journal of Psychosocial Rehabilitation* 24(7), 479–487.
- Seyhan, A.A. & Carini, C., 2019, 'Are innovation and new technologies in precision medicine paving a new era in patients centric care?', *Journal of Translational Medicine* 17(1), 1–28. <https://doi.org/10.1186/s12967-019-1864-9>
- Sivarajah, U., Kamal, M.M., Irani, Z. & Weerakkody, V., 2017, 'Critical analysis of big data challenges and analytical methods', *Journal of Business Research* 70(2017), 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Sovacool, B.K., Axsen, J. & Sorrell, S., 2018, 'Promoting novelty, rigor, and style in energy social science: Towards codes of practice for appropriate methods and research design', *Energy Research & Social Science* 45(2018), 12–42. <https://doi.org/10.1016/j.erss.2018.07.007>
- Sterner, B. & Franz, N.M., 2017, 'Taxonomy for humans or computers? Cognitive pragmatics for big data', *Biological Theory* 12(2), 99–111. <https://doi.org/10.1007/s13752-017-0259-5>
- Sun, Z., Strang, K. & Li, R., 2018, 'Big data with ten big characteristics', in *Proceedings of the 2nd international conference on big data research*, pp. 56–61, ACM, Weihai, October 27–29, 2018.
- Szopinski, D., Schoormann, T. & Kundisch, D., 2019, 'Because your taxonomy is worth IT: Towards a framework for taxonomy evaluation', in *27th European conference on information systems*, ECIS, Stockholm, June 08–14, 2019.
- Taherdoost, H., 2022, 'Different types of data analysis; data analysis methods and techniques in research projects', *International Journal of Academic Research in Management* 9(1), 1–9.
- Uğur, N.G. & Turan, A.H., 2020, 'Understanding Big Data', in *Big Data Analytics for Sustainable Computing*, IGI Global, pp. 1–29.
- Vargas, L., 2018, 'Smart media: Museums in the new data terroir', in K. Drotner, V. Dziekan, R. Parry & K.C. Schröder (eds.), *The Routledge handbook of Museums, media and communication*, pp. 261–273, Routledge International Handbooks, Routledge, Abingdon.
- Vasile, E. & Simion, D.O., 2021, 'Methods for storing and finding data in the business logic for economic applications', *Internal Auditing & Risk Management* 16(3), 9–18.
- Vassakis, K., Petrakis, E. & Kopanakis, I., 2018, 'Big data analytics: Applications, prospects and challenges', in G. Georgios Skourletopoulos, G. Mastorakis, X.C. Mavromoustakis, C. Dobre & E. Pallis (eds.), *Mobile big data*, pp. 2–20, Springer, Cham.
- Vecchi, E., Pospíšil, L., Albrecht, S., O'Kane, T.J. & Horenko, I., 2022, 'eSPA+: Scalable entropy-optimal machine learning classification for small data problems', *Neural Computation* 34(5), 1220–1255. https://doi.org/10.1162/neco_a_01490
- Wang, L., 2017, 'Heterogeneous data and big data analytics', *Automatic Control and Information Sciences* 3(1), 8–15. <https://doi.org/10.12691/acis-3-1-3>
- Wang, J., Yang, Y., Wang, T., Sherratt, R.S. & Zhang, J., 2020, 'Big data service architecture: A survey', *Journal of Internet Technology* 21(2), 393–405.