ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

TRANSFORMING HEALTHCARE WITH BIG DATA ANALYTICS

Archana Chandrakant Patil
Assistant Professor, Department of Information Technology, SIES (Nerul) College of Arts, Science& Commerce
archanapatil@sies.edu.in

ABSTRACT

Big data analytics is a developing field that could offer valuable information for the healthcare industry. While managing the volume, variety, velocity, veracity, and value of big data still presents challenges in its usage and acceptance, accuracy, integrity, and semantic interpretation are of more significance in healthcare applications. However, despite these difficulties, big data is still being used and explored in the healthcare industry as a source of evidence. This fuels the need to research healthcare information to manage and contain the skyrocketing expense of healthcare and to look for proof to enhance patient outcomes. While there are several well-publicized examples of the use of big data in health, such as Google Flu and HealthMap, there is no one, comprehensive method for doing so.

Keywords: *Literature review, systematic review, big data analytics, healthcare.*

INTRODUCTION

About 90% of the data in use now, according to a Commonwealth of Australia assessment from 2013, was produced in the previous two years. Data production will increase 44 times from 2009 to 2020, according to calculations. According to other figures, 2.5 quintillion bytes of data are generated per day. The following definition has been established for the study's purposes: "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, novel forms of information processing for improved insight, decision making, and process optimization" (Commonwealth of Australia, 2013, p. 8). This definition acknowledges big data's three dimensions. Veracity and worth are arguably two additional difficulties.

Big data is understood to be a transdisciplinary system for processing information. Big data is being increasingly incorporated into information processing systems across various commercial, government, media, and industry sectors, particularly healthcare. Understanding the components of the 2.5 quintillion bytes of data, where they are stored, whether they are raw, processed, or derived artifacts, and how public and private access is defined is necessary to fully utilize the promise of big data in healthcare. Right now, there are no solutions to these queries. To begin responding to such a question, categorize the regions of current use to create a meaningful picture.

To study how big data may be applied to specific areas to achieve the most advantage for the targeted research, the initial review aims to establish a framework for understanding the use of big data in healthcare. Big data analytics is relevant to healthcare because it offers a way to address the challenges of its information systems. Despite being categorized separately, the five big data aspects interact.

- The management of data volume, which is typically expressed in terms of terabytes or petabytes, is referred to as a data scale. It entails managing data storage (Feldman, 2012).
- Variety (a range of data types): Data formats include structured, semi-structured, and unstructured formats (Feldman, 2012)
- The rate at which data is produced, processed, and analyzed is referred to as velocity (Feldman, 2012).

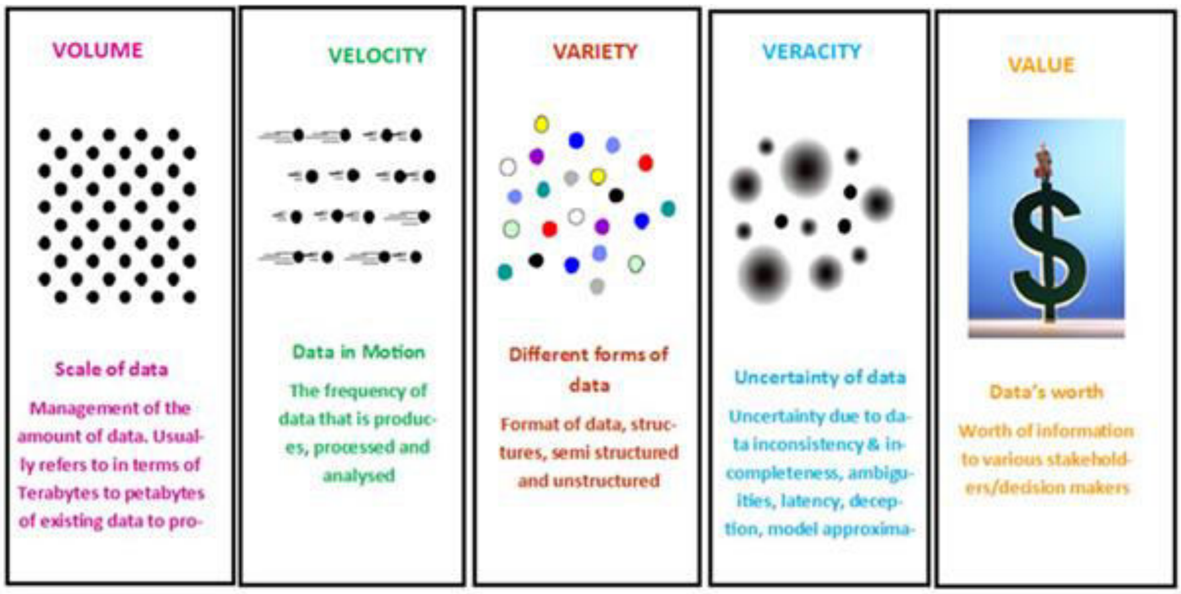
ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

- Veracity (data uncertainty): The caliber, relevance, predictive power, and significance of the data (Clifford, 2008).
- Information's value to different stakeholders and decision-makers (Clifford, 2008).

Figure 3 Big data use in healthcare categorization

The five big data dimensions are shown in Figure 1. (adapted from Haas, 2013)



These five factors are crucial in addressing the problems with big data analytics. Big data analytics is the process of analyzing large volumes of data from both structured and unstructured sources (Commonwealth of Australia, 2013). Additionally, it relates to the number of data sets used in the data analysis as well as the speed at which those data sets are analyzed.

As healthcare systems have become more complex and expensive, big data has recently been introduced to the industry as a solution to many information system issues relating to healthcare (Sun, 2013). According to estimates, the amount of healthcare data in 2012 was close to 500 petabytes. While predictions for the future indicate that by 2020, healthcare data would amount to 25,000 petabytes. An effective blending of Big data is the brains behind Electronic Health Records (EHRs), as it can link operational, clinical, and financial systems and enable evidence-based medicine. A systematic analysis of prior clinical data is done in the context of evidence-based medicine to inform decision-makers. There is evidence that big data can be used to diagnose illness, and can especially help with clinical genomic analyses of HIV patients (Feldman, 2012). However, to achieve the objectives of big data analytics, large data requires rigorous data management (Commonwealth of Australia, 2013). This covers data stewardship, user training, data access and security, data content, quality, consistency, and governance of the data sources (Shaw, 2013). Data-related problems may occur in the absence of appropriate management.

While addressing these challenges is crucial, it is also crucial to explore the real-world use and possibilities of this novel phenomenon by looking at the landscape of trial use and leading applications in big data. This kind of an examination has not yet been documented in the literature. Identifying and classifying the current cutting-edge uses is the first step in promoting effective big data use. To categorize the uses of big data in healthcare, this research offers a comprehensive review of big data. The method of research is described, along with an explanation of why certain papers were chosen for the review. The analysis of the articles and the resulting categorization of the results are then presented.

METHODOLOGY

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

A systematic review's objective is to summarise the literature on a certain topic using data from prior investigations (Clarke, 2011). As with any research project, a comprehensive systematic review must adhere to a predefined plan. It is a methodical approach to reviewing the literature on a particular topic that, like any primary research, is well-documented to enable replication. It entails formulating a review or research question, assessing the body of available evidence and developing pertinent keywords, conducting a thorough literature search, developing inclusion and exclusion criteria, synthesizing the findings from various studies, conducting analysis, and reporting the findings (JBIEBNM, 2000).

Formulation of Research/Review Question

The methodology's first phase defined a suitable question. Can large data be classified into discrete usage categories to reduce semantic misinterpretation, resulting in better big data analytics performance in healthcare? was the main research topic.

- What applications of big data do there exist in the field of healthcare?
- How is the use of big data categorized?

Examining the Evidence Base, Creating Relevant Keywords, and Developing Selection Criteria

To find prospective sources of literature, as well as terminology and keywords linked to the use of big data in healthcare, an initial review of the literature, was done. The research's original focus was only on works published within the previous 14 years. This was brought on by big data and health analytics' relatively recent evolution. As a result, publications that were less than three years old were more valuable than those that were released more than fourteen years ago. Reliable sources conducted high-caliber research that was approved for usage. Blogs and other internet-based publications were not considered to be of high scientific quality. Finding relevant literature required exploring online academic resources like JAMIA, Wiley Online, and

The second phase involved extending the initial search terms and looking for relevant literature that matched the requirements. 30 publications were chosen from the first literature search as having potential for inclusion in the study. To conduct a comprehensive literature search process, search terms were established from the original literature evaluation. Each item was classified according to search terms, database source, and the number of articles discovered (Table 1). (Figure 2).

Table 1: Number of publications located using final search terms

Search Terms Used	Number of Publications
Big data and clinical decision support	7
Big data and healthcare delivery	6
Big data and health services	4
Big data and healthcare admin	0
Big data consumers	4
Big data patients	2
Big data behavior	3
Big data clinical data management	4
Big data healthcare data management	1
Big data and Electronic Health Records	4
Big data and public health	5
Total number of publications	40

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022**

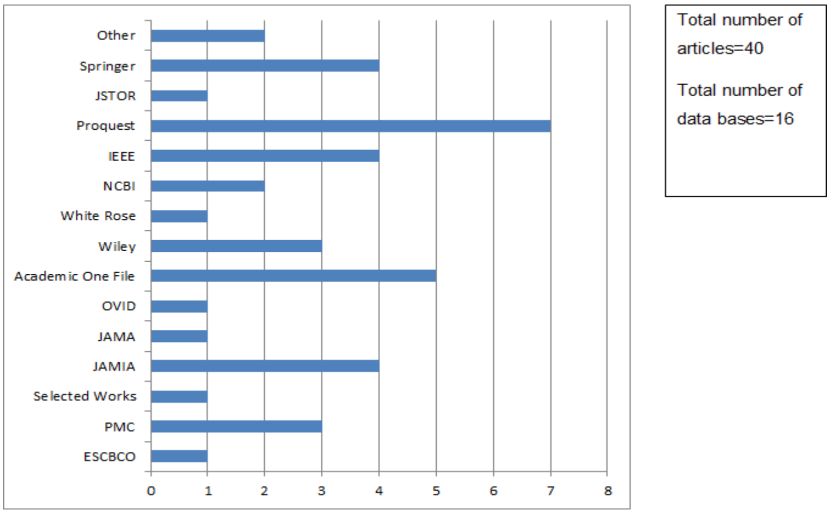


Figure 2 Databases and number of articles identified

40 articles were found after using the aforementioned terms to search the literature. The term "big data and healthcare administration" did not turn up any pertinent articles. This does not imply that there is no study on big data and healthcare administration, but it does imply that there are few results when searching using that specific research keyword. Out of the 40 articles, seven were from ProQuest and five were from JAMIA. The use of big data in healthcare is a young and developing topic that frequently lacks classification in studies. It is possible to examine what is currently being published regarding the use of big data in healthcare by looking at the sources of the articles.

Combining the findings of many studies

The third step entailed sorting through the literature search findings. The collected literature was divided into low-quality research literature and high-quality useful material. High-quality content is defined as papers from peer-reviewed academic articles that included pertinent data. It was able to find the articles that were pertinent to the research issue by reading and analyzing the literature. Given the variety of contexts, the four broad categories listed below—proposed by Groves (2013)—were utilized to offer the first categorization and to enable preliminary analysis.

1. Managing healthcare delivery costs and administration: The management of healthcare services is referred to as administration and delivery. Numerous tasks that fall within the categories of administration and delivery can be demonstrated as being used by big data analytics. The use is classified in this category when it has an impact on how healthcare delivery is managed.
2. Clinical decision support: Clinical decision-making can be supported by big data. Clinical decision support encompasses a variety of elements that give physicians access to additional data to help in decision-making.
3. Clinical data: The data and information set available expressly for big data analytics are represented by this category. Within this category, it is clear that there are several kinds of clinical information systems as well as information systems that store clinical data.
4. Consumer/behavior: Within the consumer/behavior category, big data used for demographic analysis are acknowledged. Here, both individual and societal behavior and lifestyle elements are included.

As the analysis went on, it became clear that a fifth category was needed to account for useful articles and annotations that didn't fall under one of the main broad Groves categories. This group was titled "support information."

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

5. Supporting Information: Uses of big data that are not categorized into one of the other four categories fall under this category. Its use either falls under more than one category or has been demonstrated to fall under a different category all on its own.

ANALYSIS AND RESULTS

The articles described a wide range of potential uses. As extrapolated from Groves, Table 2 shows the numerous utilisations based on the extended generic categories (2013)

Table 2: Analysis of articles using Groves (2013) extended categories

Groves Extended Categories	Article Reference	Type of Information Located
Administration and delivery: managing healthcare delivery costs	Murdoch & Detsky, 2013	Electronic Health Records, decision-making, observational-based evidence, patient data, trials, case studies, quantitative data, real-time patient data, genomics demographics, lifestyle information from other sources, healthcare delivery, economic values, medication lists, family history, practitioner data, knowledge dissemination, direct information to patients
	Harper, 2013	Predict public health, patient care, and population health, Clinical decision support: automated algorithms, electronic health record secondary data. business data, economic data, billing data, financial data, reduced healthcare costs, education
	Rose & Burgin, 2014	Monitoring of evidence-based guidelines for prevention screening, treatment plan, Electronic Health Records -population data, patient-specific information, patient-reported data, physician-collected information, medication records, patient behavior, reduce costs, predictive analysis-risk accuracy, medication adherence
	Monheit & Berk, 2001	Health care costs, health insurance
	Sheldon, 1998	Organizational healthcare, assurance healthcare, state medicine, quality indicators,
	Ferlie & Shortell, 2001	Track population, disease registry, clinical information -outcome data, evidence-based, financial data, process data, legal data, education, academic, guidelines
Clinical decision support	Adler-Milstein & Jha, 2013	Early detection of treatment, lower costs, Clinical decision support, Electronic Health records-Bio surveillance, Accurate predictions of sick, health behaviors
	Bates et al., 2003	Health information systems, Evidence-based medicine, clinical decision support, Electronic Health Records
	Weber et al., 2009	Ontology, health observations, demographics, Clinical data, health research information, patient information
	Thadani, Weng, Bigger, Ennever, & Wajngurt, 2009	Electronic Health Records, clinical trials, semantic dictionary, Predictive analysis, E- screening, demographics
	Swinglehurst, Pierce, &	Decision-making, GP evidence-based medicine, primary care delivery, clinical information,

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

	Fuller, 2001	Local health, community health
	Peleg & Tu, 2006	Clinical data, standard terminology, healthcare delivery, patient care, reducing costs- prevention of medication errors, educational institutions
	Cooper, Kennedy, & Springer, 2007	Healthcare management-healthcare services, laboratory results, healthcare databases
	Chen, Mao, & Liu, 2014	Public health, biomedical data, R and D, genomics, clinical gene diagnosis fraud claims, error investigation, tax claims
	Cushing, 2013	Visual analytics, research
	Howe et al., 2008	Biological data, research, semantics
Clinical information	Liu & Park, 2014	External health data (social media), Electronic Health Records, mobilized Health Records, health monitors, genetic sequencing, biomedical sensors
	Doug et al., 2008)	Confidential data-machine learning-clinical trials, epidemiological database, research
	Schouten, 2013	Predictive modeling, anomaly detection, pattern detection
	Sanat, 2013)	Real-time consumer sentiment, social interaction, financial data, predictive diagnostic, warranty analysis, genomic research
	Chawla & Davis, 2013	Decision-making, disease risk profiles, disease prevention, disease management, clinical data, population-based evidence healthcare, individual-based evidence healthcare, genomics, lifestyle, and environmental behaviors to contribute to factors
	Kuriyan & Cobb, 2013	Reduce costs, insurance claims, health claims data, demographics: social determinants- health population, prescription data-forecasting chronic disease
	Eramo, 2013	Electronic health record clinical data, disease registry data
	Skolnik & Bertman, 2013	Electronic health records, population chronic disease management, physician behavior, treatment management, clinical decision support, evidence-based medicine, diagnosis and therapeutic interventions, Electronic health record data analytics
	Kevin, 2008	Electronic health records
	Bonney, 2013	Healthcare reform, electronic health records, population healthcare
	Ackerman, 2012	Genomics disease treatment, outcomes
	Green, 2006	Clinical care, disease epidemiology, social behavior
	Diehr & Lumley, 2002	Predictive public health, public health
	Ringel & Skiera, 2014	Consumers, competition
	Kaye, 2013	Marketing, consumer behavior
	Wang & Strong, 1996	Advertising, customers, behavior
	Timothy & Susan, 2002	Healthcare consumer, demographics, external healthcare tracking
	Kallus, 2014	Predictive analysis within the population
	Bentley, O'Brien, & Brock, 2014	Decision-making, health behavior, social phenomenon, public health, research
	Yoon, Park, Kim,	The behavior of individuals and public

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022**

	ChaeTae & JunHyun, 2014	
	Gross & Bates, 2007	Clinical trials, clinical information, clinical laboratory, cost-effective, detection of errors, medications, clinical knowledge for education-clinical decision support, evidence-based medicine
	Ola & Sedig, 2014	Decision-making, policy visual analytics, analytic reasoning, interpretation, problem-solving, education
	Hoffman & Podgurski, 2013	Clinical care policy, immunizations, treatment effectiveness, electronic health records, clinical outcomes, public health research, surveillance of infectious disease, outbreaks, lab results, research, public health interventions

There are many different ways to use big data in healthcare, as seen in Table 2 even from these 40 papers. Following is a list of these conclusions:

- In seven of the forty papers, it was proposed that big data analytics may be utilized to forecast some sort of disease or condition.
- Seven people talked about a cost-cutting possibility.
- Semantic standards in big data analytics were discussed in three articles. The literature analysis also revealed a range of additional uses, including primary care, health care delivery, and others.

Regarding the subcategories, a trend was discovered; these either directly or indirectly influenced healthcare delivery management. To manage healthcare delivery costs, it was decided that each utilization that involved management operations should fall within the administration and delivery category.

- Twelve publications made the case that big data analytics would incorporate electronic health records.
- Five people said that an important result of big data analytics would be evidence-based medicine. Further help for clinical decision support was produced by other big data initiatives in the healthcare industry.
- Ten papers make the case that clinical decision support will result from big data analytics in the healthcare industry.

The literature suggests that big data analytics may process other data, including electronic health records, and produce valuable information that can be utilized to enhance clinical decision-making and evidence-based medicine.

- A variety of data and information were mentioned in sixteen publications as being used in big data healthcare analytics.
- Big data analytics' importance in genomics was discussed in five papers.
- A big data role in public health was represented by fifteen papers.

The big data healthcare activities reported in the literature showed a pattern of databases and data that big data analytics may employ to map out possible disease outbreaks in the general population. Due to the mapping of individual data, this would be conceivable. Given that they entail the many available kinds of clinical information, these activities were determined to be appropriate for the category of clinical information. genomes, patient data, and clinical trial information, among other things.

- The utilization category behavior/consumer was illustrated in fifteen articles from the literature review.

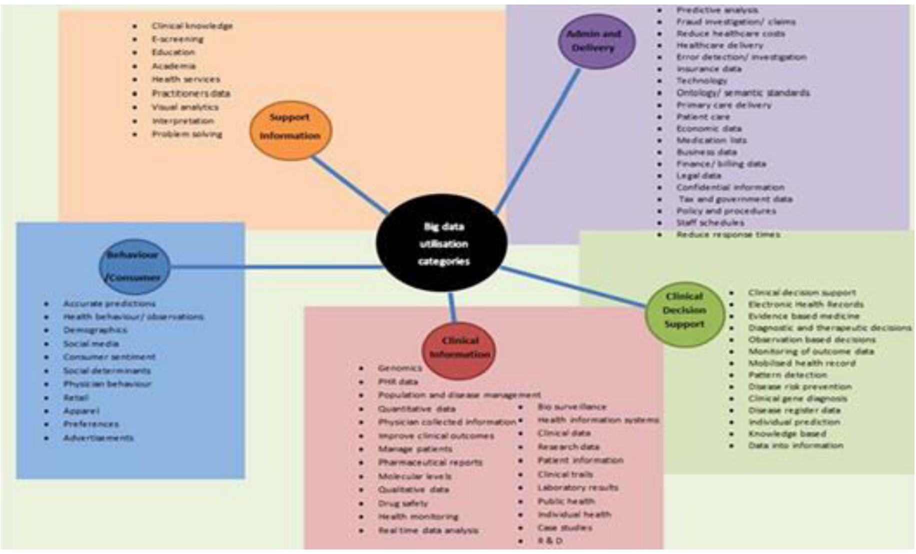
ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

It was highlighted that there is literature on patient lifestyles, health behaviors, demographics, and observations of health. It has been demonstrated that big data analytics can follow and map people's lifestyles, which enables the prediction of both people's behavior and that of the general public. A pattern involving the fusion of all the categories is therefore obvious.

Support data and big data usage that didn't fit the established categories were put in other. It should be noted that big data analytics functions within the support information category interact with those of the other categories. It is clear that despite being separate, the big data utilization categories do interact.

Figure 3 was constructed from an analysis of the articles in Table 2 and represents the currency of big data use within healthcare.



CONCLUSION AND FUTURE RESEARCH

Without a doubt, effective data mining and medical informatics integration and its analysis utilizing big data approaches will have an impact on the price of healthcare delivery and lead to better healthcare outcomes through well-informed decision-making (Sun, 2013). A cross-section of different papers was taken for use in the study from a systematic review of the literature. Multiple databases contained the literature, indicating that there is still no clear standard or natural location for publishing big data in healthcare. Additionally, the scant number of academic articles indicates that, although health informatics and data science are a part of healthcare, it has not yet developed a strong academic foundation and body of literature. Despite this, there are many marketing cases and proposed

Utilization categories were developed from the literature's comprehensive review. Semantics are included in the categories. The development of categories may lead to the development of language and standard phrases for use in big data analytics. Indeed, this could lead to increased yield and more focused searches to support evidence-based medicine and create trustworthy and effective decision support for patient diagnosis and care. Finding out what literature about big data analytics in healthcare is now being published may lead to more targeted publishing that can help in identifying the current usage of big data analytics. Determining which databases will be useful for the next big data analytics study can also come from the identification of the databases.

The existing literature can be categorized into the following areas for the use of big data indicated by this study: administration and delivery, clinical decision support, clinical information, behavior/consumer, and support information. An extensive categorization of big data uses in healthcare has been mapped using an expanded classification technique to classify studies. This offers a foundation for investigating how to "tame" the use of large data to advance understanding. Additionally, this will make it possible to identify the different categories of analytical instruments

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 10, Dec 2022

that might be used later. The study paints a picture of the numerous and varied ways that big data are used in healthcare. This shows that its use today is opportunistic rather than systematic. Through these initial applications, the potential is being investigated and shows promise for quick

REFERENCES

Sheth, A. (1996). Data semantics: what, where, and how. Paper presented at the 6th IFIP Working Conference on Data Semantics (DS-6), Atlanta, GA.

Ackerman, M. (2012). Big data. The Journal of Medical Practice Management: MPM, 28(2), 153.

Adler-Milstein, J., & Jha, A. K. (2013). Healthcare's "Big Data" challenge. American Journal of Managed Care, 19(7), 537-538.

Clarke, J. (2011). What is a systematic review? Evidence-Based Nursing, 14(3), 64. DOI: 10.1136/ebn.2011.0049 Clifford, L. (2008). How do your data grow? Nature, 455(7209), 28.

Commonwealth of Australia. (2013). The Australian public service big data strategy: Improved understanding through enhanced data analytics capability. Canberra: Commonwealth of Australia.

Cooper, R., Kennedy, J., & Springer, V. (2007). Data management: data, data everywhere, In Proceedings of the 24th British National Conference on Databases, BNCOD 24, Glasgow, UK, July 3-5, 2007: proceedings (Vol. 4587.). New York: Springer Berlin Heidelberg.

Cushing, J. B. (2013). Beyond big data? Computing in Science & Engineering, 15(5), 4-5. doi: 10.1109/mcse.2013.102