Chapter 3
Big Data in Higher Education: The Big Picture

Ben K. Daniel

Abstract Globally, the landscape of higher education sector is under increasing pressure to transform its operational and governing structure; to accommodate new economic, social and cultural agendas; relevant to regional, national and international demands. As a result, universities are constantly searching for actionable insights from data, to generate strategies they can use to meet these new demands. Big Data and analytics have the potential to enable institutions to thoroughly examine their present challenges, identify ways to address them as well as predict possible future outcomes. However, because Big Data is a new phenomenon in higher education, its conceptual relevance, as well as the opportunities and limitations it might bring, is still unknown. This chapter describes the conceptual underpinning of Big Data research and presents possible opportunities as well as limitations associated with unlocking the value of Big Data in higher education.

Keywords Big Data • Learning analytics • Higher education

Introduction

Globally, the landscape of higher education sector is under increasing pressure to transform its operational and governing structure; to accommodate new economic, social and cultural agendas; relevant to regional, national and international demands. As a result, universities are constantly searching for actionable insights from their data, to generate strategies they can use to meet these new demands. Among many others, the pressure to implement new far-reaching changes in higher education is triggered by individual and combined forces of globalisation. The direct influence of globalisation in higher education can be seen in the neoliberal reform initiatives at universities, including increase in the number of students, massification and marketisation of higher education, continuous reduction in funding to support

B.K. Daniel (✉)
Educational Technology Group, Higher Education Development Centre, University of Otago, Dunedin, New Zealand
e-mail: ben.daniel@otago.ac.nz

B. Kei Daniel (ed.), *Big Data and Learning Analytics in Higher Education*,
DOI 10.1007/978-3-319-06520-5_3

Big Data is regarded as term that describes an incredible growth in volume, structure and speed in which data is being generated. Douglas (2001) provides a summary of what constitutes Big Data in what has commonly come to be known as the "three Vs" (volume, velocity and variety) as a way of understanding the structural features of what can regarded as Big Data. Generally, the literature extends from the three-core features and presents the following as key characteristics associated with the notion of Big Data (Daniel, 2015):

- *Volume*—used to describe a large amount of information that is often challenging to store, process, transfer, analyse and present.
- *Velocity*—term associated with increasing rate at which information flows within an organisation—(e.g. institutions dealing with financial information and relating that to human resources and productivity).
- *Veracity*—refers to the biases, noise and abnormality in data generated from various sources within an organisation. Veracity also covers questions of trust and uncertainty associated with the collection, processing and utilisation of data.
- *Variety*—referring to data presented in diverse format both structured and unstructured.
- *Verification*—refers to data corroboration and security.
- *Value*—refers to the ability of data in generating useful insights, benefits, business processes, etc., within an institution.

There are also other important features of Big Data (see Fig. 3.1) such as data validity, which refers to accuracy of data, and data volatility, a concept associated with the longevity of data and their relevance to the outcomes of the analytics. It also refers to the length and time required to store data in a useful form for further appropriate value-added analysis.

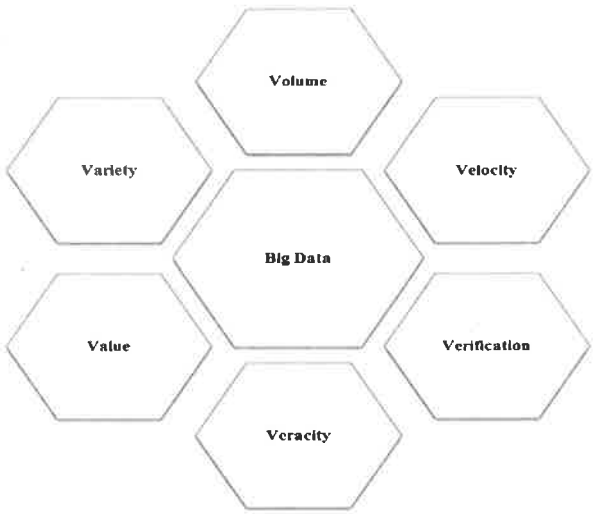

Fig. 3.1 Key characteristics of Big Data

Examples of current implementation include experimentations with early detection systems (EDS) (Dawson, Bakharia, & Heathcote, 2010; Siemens, 2013); using the information provided by an EDS, faculty can learn about students and their learning environments and provide needed targeted interventions.

Learning analytics is concerned with the measurement, collection and analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs (Siemens & Long, 2011). Learning analytics is undertaken more at the teaching and learning level of an institution and is largely concerned with improving learner success (Jones, 2012). For instance, learning analytics can be used to understand students' behaviours in learning management systems.

The widespread introduction of learning management systems (LMS) such as Blackboard and Moodle resulted into increasingly large sets of data. Each day, LMS accumulate increasing amounts of students' interaction data, personal data, systems information and academic information (Romero et al., 2008). LMS keep record of students' key actions. Student data in LMS could also include more detailed information on the content of students' postings or other writing, their choices and progress through a specific interactive unit or assignment or their particular preferences and habits as manifested over a range of tasks and interactions or semester (Friesen, 2013). Using learning analytics, this information can be used to understand student behaviour, the learning environment, teaching effectiveness as well as the environment in which teaching is performed.

Opportunities and Limitations

The use of Big Data can inform the next innovation in higher education (Siemens (2011) and advance educational research. For a number of years, researchers in higher education have worked with relatively small amounts of data to research their disciplines. This is because the tools for collecting, organising, analysing and presenting data are limited. They have also relied on methods such as perception-based data gathered through surveys or interviews that have relatively limited interpretative power and latency validity.

In conventional research design, researchers depend on probability and non-probability sampling techniques to guide them in capturing data, which can be costly and limiting in interpretation (Mayer-Schönberger & Cukier, 2013). Big Data methodologies enable researchers to work with a large set of data, removing the barriers of sampling inadequacy, increasing generalisability confidence and enhancing validity rigour. In addition, by working with large amount of data, researchers can use advanced statistical clustering techniques to investigate in more details a particular subgroups within a population without necessarily relaying on expensive probabilistic techniques.

Additionally, when used effectively, Big Data can help institutions enhance learning experience and improve student performance across the board, reduce

ity to support different forms of privacy and data security. It is also worth noting that presently the field of Big Data and analytics is dominated almost entirely by technology professionals, many of whom have limited pedagogical knowledge needed to effectively support learning. This might change in the future with the establishment of Data Science programmes that provide learning technologies as areas of specialisations.

The collection of educational data raises issues on ethics associated with data ownership, privacy, security and ethics of use (Jones, 2012; Prinsloo et al., 2015). There are also matters of accountability associated with the use of student data for predictive modelling. For instance, Eynon (2013) noted that the more we know about student challenges during their education, it is likely that the social implications of decisions to support learning will become more complex. For instance, it becomes difficult to make a decision on what to do when we know that a student is more likely to drop out of their programmes.

Further, research driven by Big Data is by large limited to correlational models and predictive analytics. In others words, through the use of data mining techniques, researchers search through an ocean of data but only answer questions of "what" rather than "why". Outcome of educational research is often needed to address learning problems; while asking "what" questions are necessary, it is not sufficient to provide necessary advice to support educational outcomes. Additionally, the utilisation of Big Data has hidden biases inherent in both the collection, analysis and reporting. Reliance on the results of Big Data alone without other sources of evidence, such as experience, can be misleading and likely to disadvantage some individuals or subgroup of people in an institution.

Conclusion and Future Research

The reliance on data-driven decision-making will become a central approach in many research- and teaching-intensive institutions in higher education. Big Data and analytics are more likely to effectively transform the way higher education operates and governed itself through the use of various technologies to capture, process, analyse, present and use data to generate actionable insights to drive their decisions.

Key drivers of the implementation of Big Data and analytics in higher education can be linked to increasing pressure to base evidence on data rather than intuition or experience. Other factors include increased accountability demanded by stakeholders, which in turn, necessitates the collection of different forms of data for the purpose of generating reports for internal and external regulations. The increasing use of various information technologies by students, faculty and other staff generates a vast amount of data that can be mined for useful information. The emergence of unstructured forms of new data often associated with social media technologies (images, tweets, videos, audios and web pages) and the value that might accrue from processing this data (Sagiroglu & Sinanc, 2013) in higher education is an additional reason for exploring data to gain useful information.