<table>
<thead>
<tr>
<th><strong>Metadata of the chapter that will be visualized online</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter Title</strong></td>
</tr>
<tr>
<td><strong>Copyright Year</strong></td>
</tr>
<tr>
<td><strong>Copyright Holder</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td>Family Name</td>
</tr>
<tr>
<td>Particle</td>
</tr>
<tr>
<td>Given Name</td>
</tr>
<tr>
<td>Suffix</td>
</tr>
<tr>
<td>Organization/University</td>
</tr>
<tr>
<td>City</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td><strong>Keywords</strong> (separated by “-“)</td>
</tr>
</tbody>
</table>
**Education and Big Data**

Maggi Savin-Baden

University of Worcester, Worcester, UK

Introduction

Big data is a big concern for many academics, largely because it is complex, unmanageable, and open to misuse. While there is a tendency to believe that “big data” might be bad and possibly dangerous, many types and uses for it exist. The challenge of big data for higher education is that it has been, until fairly recently, portrayed as something that is straightforward, clear, and easily delineated, when in fact it is none of these, and there is still relatively little consensus about how it might be defined. This entry explores how big data is defined, described, and utilized in different contexts. It explores different notions of analytics and suggests how these are having an impact on higher education. The entry then explores the claims that are being made about the objectivity of big data and sets these claims in the broader context of what can be claimed and what cannot. In the context of such claims, the way in which the ideas about what is plausible, possible, and honest in the use of big data is examined, and suggestions are offered about what may be useful and realistic uses of big data. The final section explores the possible futures for research and use of big data in the context of higher education and offers some suggestions as to ways forward.

Definitions of Big Data

For many researchers in higher education and across the disciplines in general, big data is invariably expected to offer new insights into diverse areas from terrorism to climate change. Yet at the same time, big data is also troublesome, since it is perceived to invade privacy and increase control and surveillance. Definitions of big data are wide and varied, for instance, there are definitions that concentrate on scale or diversity and others that focus on the economics of big data. For example, Taylor et al. (2014, p. 3) cite examples such as the number of variables per observation, the number of observations, or both, given the accessibility of more and more data – what Varian, Chief Economist at Google, referred to as “fat data, long data, extensible data, and cheap data.” However, Einav and Levin (2013) argue for three main features of big data:

1. Sources are usually available in real time.
2. The scale of the data makes analysis more powerful and potentially more accurate.
3. Data often involve human behaviors that have previously been difficult to observe.
Kitchin (2014, pp. 1–2) delineates big data as:

- Huge in volume, consisting of terabytes or petabytes of data
- High in velocity, being created in or near real time
- Diverse in variety, being structured and unstructured in nature
- Exhaustive in scope, striving to capture entire populations or systems (n = all)
- Fine grained in resolution and uniquely indexical in identification
- Relational in nature, containing common fields that enable the conjoining of different data sets
- Flexible, holding the traits of extensionality in that it is possible to add new fields easily, as well as expand in size

There is little consensus about what counts as big data, but many across the higher education sector see it as worthy of attention. Conceptions of big data tend to fuse across the realms of collecting large data sets and the processes of managing such data sets as well as examining how, by whom, and for whom the data sets might be used. For scientists, Kitchin’s stance (Kitchin 2014) seems a good fit, but those in social sciences and humanities tend to use the term data differently. For example, researchers in the social sciences see big data encompassing not just large date sets but also the complexity of how data are synthesized, the ways in which tools are used, and who makes which decisions about management of possible imbalances between data collection, management, and synthesis. Sometimes, assumptions and uses related to big data can be naive (Brynjolfsson and McAfee 2012). Further, there are a number of difficulties with big data analysis such as the shortcomings of off-the-shelf packages, the storage of data, and possible efficiencies in distributed processing. Table 1 summarizes different ways in which disciplines are seeing and using big data.

**Assorted Analytics**

There are currently many different types of analytics in higher education, but it is only relatively recently that it has been termed learning analytics. However, learning analytics is in fact rooted in a longer tradition such as educational data mining and academic analysis. Currently (in 2015), learning analytics in education and educational research focuses on the process of learning (measurement, collection, analysis, and reporting of data about learners and their contexts), while academic analytics reflects the role of data analysis at an institutional level. For many researchers in higher education, learning analytics and data analytics are seen as fields that draw on research, methods, and techniques from numerous disciplines ranging from learning sciences to psychology.

This melange of ideas, constructs, and approaches is reflected in the varieties of methodologies being used across different institutions. For example, in the process of analyzing big data, discipline-based pedagogy and disciplinary difference are often transposed in ways that do not necessarily reflect the nuances of the discipline. Furthermore, it is evident that different institutions are using different approaches to collecting and analyzing data. These include Oracle data warehouse and business intelligence software, the use of QlikView to analyze data held in Microsoft SQL Server, and also Google Analytics, Google Charts, and Tableau (Sclater 2014).

While there have been various attempts to classify analytics into a clear typology, Table 2 illustrates that issues in higher education are murky and complex. Thus, it is possible to see multiple and overlapping types, including (big) data analytics, text analytics, web analytics, network analytics, and mobile analytics.

**Objectivity and Context**

There have been suggestions that big data and analytics are necessarily objective. However, the complexity of their use in different disciplines means that there is little unity about how these
data should be analyzed and used. It seems for
many researchers, particularly in areas such as
economics, that the focus is on complex analysis
of big data, rather than asking critical questions
about whether big data is new and what can and
cannot be done with it. The result is that across
the literature there is a wide range of positive and
negative claims, which need to be acknowledged,
including but not limited to:

Claim 1: Big data speaks for itself. This is clearly
not the case since analysis and mapping are
researcher driven. There is a need to ask not
just what might be done with big data but why
(and if) it should be used in particular
ways – as well how big and small data might be used together.

Claim 2: There are many good exemplars of big
data use. This is not the case, particularly in
the social sciences and education, where the
landscape is complex and varied. For exam-
ple, in 2013 Snowden disclosed that the US
National Security Agency was monitoring
domestic “metadata.” The archive released
by Snowden indicated that the e-mails, phone
calls, text messages, and social media activity
of millions of people around the world had
been collected and stored and then without
consent been shared and sold (Rodriguez
2013). Although this has brought to light a

### Education and Big Data, Table 1

<table>
<thead>
<tr>
<th>Context</th>
<th>Understanding and use of big data</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economics (Taylor et al.)</td>
<td>Specific terminology seen as fairly recent – some were working with what is now being termed “big data” a decade ago and believe it has not gained much traction within academic economics</td>
<td>Seen as a class of data which was particular in terms of its size and complexity, although there were several different points of view as to which features rendered it genuinely new</td>
</tr>
<tr>
<td>Digital humanities (Manovich 2012)</td>
<td>The use of data analytics to analyze and interpret cultural and social behaviors</td>
<td>Complex overview of data, visual representations of images and videos, exploration of patterns of representation</td>
</tr>
<tr>
<td>Education (Sclater 2014)</td>
<td>Use of data to analyze student retention, student engagement, and identification of risk and to examine student progress</td>
<td>Data seen as useful for gaining information, tracking possible problems by student, tutors, and senior management</td>
</tr>
<tr>
<td>Business (Brynjolfsson and McAfee 2012)</td>
<td>Use of data to make predictions and management decisions</td>
<td>Information from social networks, images, sensors, the Web, or other unstructured sources are used for decision making in business</td>
</tr>
<tr>
<td>Journalism (Lewis 2015)</td>
<td>Journalism that incorporates computation and quantification in diverse ways, for example, computer-assisted reporting</td>
<td>The implementation of mathematical skills in news work as well as the critique of such computational tools</td>
</tr>
<tr>
<td>Maths and statistics (Housley et al. 2014)</td>
<td>Creating mathematical tools for understanding and managing high-dimensional data</td>
<td>Tools, algorithms, and inference systems seen as vital for analysis of data within maths and statistics but also other disciplines using big data</td>
</tr>
<tr>
<td>Computer science (Rudin et al. 2014)</td>
<td>Use of methods for statistical inference, prediction, quantification of uncertainty, and experimental design</td>
<td>Use of multidisciplinary teams with statistical, computational, mathematical, and scientific domain expertise with a focus on turning data into knowledge</td>
</tr>
<tr>
<td>Medicine and health (Lee and Yang 2015)</td>
<td>Predicting and modeling health trends</td>
<td>Locating health patterns Understanding prevalence and spread of disease</td>
</tr>
<tr>
<td>Psychology (Moat et al. 2014)</td>
<td>Predicting and modeling trends and the use of data sets to examine behaviors, influence, and use of language</td>
<td>Analysis of trends, behaviors, judgment, and decision making as well as spheres of influence</td>
</tr>
</tbody>
</table>
number of other forms of monitoring and surveillance practices, the US Government argued that it was only “metadata.” The Snowden examples introduce questions for those who work in higher education about how data they collect and are data that collected about them and their students are used in covert ways. It would seem that increasingly government agencies are using big data in ways which focus on economic outcome results in unhelpful social, political, and cultural bias for educational activities. Such a stance would seem to indicate that there is increasingly a neoliberal agenda shaping higher education, with a growing belief in competitive individualism and the maximization of the market.

**Claim 3:** There is integration and understanding across the disciplines. While some universities have shared forums for big data, much big data remains in disciplinary silos. There is a need for greater interdisciplinarity and large teams to work together coherently.

**Claim 4:** There is a coherent view about how learning and academic analytics should be used. It is evident that institutions already seem to be finding themselves having to balance students’ expectations, privacy laws, tutors’ perspectives about learning, and the institution’s expectations about retention and attainment.

These four claims exemplify the need to consider issues of plausibility and honesties in big data research. What is often missing from claims and debates is how power is used, created, or ignored in the management and representation of big data, or where and whose voices are heard or ignored, privileged, or taken for granted.
Plausibility and Honesties

Ethical issues connected with using big data are complex and muddled. It is often assumed that just because data are public, ethical concerns can be ignored. The open, accessible, and online society has resulted in various kinds of uses of big data, one of which is tracking. For example, many people inadvertently leave tracking devices switched on their mobile phones or leave the Wi-Fi on overnight. The result is that people do not realize they are being tracked, while feeds from social networks are analyzed and visualized, personal movements tracked, and shopping behaviors noted. There is thus a serious lack of privacy, which occurs through the aggregation of users’ online activities. Companies can track and aggregate people’s data in ways previously impossible, since in the past people’s data were held in paper-based systems or company silos. Now, personal data can be mined and cross-referenced, sold, and reused, so that people are being classified by others through sharing their own data. This use of big data is often disregarded – but it is relatively easy to discover most things about most people, and blue chip companies can use such large data sets to ensure market advantage. In day-to-day life, this open but hidden knowledge is already both accepted and ignored. There have been discussions about the need for better formal regulation and changes to the way social media are designed. Yet, almost a decade after the concerns were first raised, the suggested changes are unlikely to occur, and it is difficult to decide how security might be maintained in a post-security world. Now, as time marches on, we are encountering various forms of liquid, participatory, and lateral surveillance (Savin-Baden 2015).

There are still questions about what it is possible to “know” from big data analyses and ethical challenges concerning what is done or not done with such analyses and findings. In higher education, the focus and interest in big data have resulted in many researchers rebadging their work as “big data research,” when in fact it is not. Particularly in the humanities, this has resulted in criticism of big data research. Some years later, the pertinent criticisms and concerns of many higher education researchers still seem to have resonance. Big data has changed how data are seen, how they are used, and how they are defined. Such shifts are changing how knowledge is seen and managed in higher education. At the cusp of higher education and commerce, big data tend to be located as neutral, objective, and reliable. This, in turn, obscures the ways in which big data are covertly managed and used and the ways in which people become constructed by and through big data.

Big data, as aforementioned, can be linked to neoliberal capitalism, and engaging the current performative enterprise practices has shifted the focus in higher education increasingly toward consumerism, the marketization of values, and the oppression of freedom. Thus, criticality and questioning are being submerged in the quest for fast money and solid learning. In areas of higher education that reject neoliberal capitalism, there is a tendency to shift away from the idea of big data as a resource to be consumed and as a force to be controlled and instead to ask questions summarized below.

How Accurate or Objective Is Big Data?

As a result of the way big data are constructed and used to make policy decisions, it is vital to recognize that these data can easily be a victim of distortion, bias, and misinterpretation. Driscoll and Walker (2014) illustrate how data access and technological infrastructure can affect research results. For example, they demonstrate how differences in timing or network connectivity can result in different results for the same experiment.

Is Big Data Better Data?

While it would be easy to suggest a binary relationship here, it is important to note that big data is not always representative, nor is it necessarily presenting a complete picture of the issues, nor may it meet high enough standards of rigor and quality. Administrators, faculty, and managers in higher education may find the promise of big data alluring. Assumptions that big data is objective, with clear outcomes that will improve retention,
increase student numbers, and ensure there is more money in the university coffers, make this promise highly seductive. Yet these data are not necessarily reliable, and using them for monetary ends in a sociopolitical system such as higher education brings high risk.

How Are Issues of Context To Be Dealt With in Big Data Research?
Big data sets need to be located contextually and there is a need to understand how big data are being used and understood and what is being claimed for them. There is a need for more robust studies and examples across higher education to provide an examination of issues of context by defining and critiquing how big data and definitions of it have changed over time.

What Are the Ethical Issues Associated with Big Data?
The ethical questions relate not only to how data are obtained, as in the Snowden affair (Rodriguez 2013), but who and what is subject to analysis. Ethical considerations also extend to how and where data are reported. For example, Eubanks (2014) researched the electronic benefit transfer card and food stamp use in the United States and suggested that those in poverty are already “in the surveillance future.” The result is that the poor and marginalized that are more easily tracked are already being judged and assumptions made about them, which may or may not be just.

To What Extent Is Big Data Creating Digital Divides?
It seems that the expense of gaining access to big data has resulted in a restricted access to this data, with higher education necessary being marginalized as a sector. Yet the abovementioned study by Eubanks (2014), as well as other studies concerning surveillance, illustrates not merely digital divides but also suggest surveillance divides.

Big Futures
Big data is useful, yet multifaceted, and offers few, if any, quick fixes for new fields of research or data management. In education, social sciences, and humanities, it would seem that relatively few researchers are engaged in analyzing massive data sets. Perhaps the most important considerations in future big data research are to:

- See big data as part of a repertoire of data collection and analytical options.
- Use big data as a means of locating areas that can or need to be explored on a smaller scale.
- Use big data in multi-methodological ways so that the research undertaken is both wide and deep.
- Recognize the advantages, disadvantages, challenges, and power issues of working with large data sets alongside small, fine-grained data.
- Acknowledge that large real-time data sets, such as those produced by social networks, often do not provide a clear or representative picture of realities.
- Recognize that full documentation of how big data were collected will probably be unavailable, and therefore the validity of such data is likely to be unpredictable and tenuous.

Conclusion
There is a prominent expectation that big data can and will deliver more than is really possible and that its questionably clear outcomes will necessarily make a difference to the complexity of human life and experience. The contrasting view questions whether big data can offer anything particularly new or innovative while being concerned about the management of big data and how they are being used in persuasive and pernicious ways. What appears to be a consistent message is that big data is difficult to manage, analyze, and evaluate. Therefore, it is uncertain how robust findings and assumptions, as well as what has been learned, might in fact be. It is vital
to recognize that big data is neither good nor bad but useful in different ways if collated and presented with honesty and plausibility at its core.

References


## Author Queries

*Encyclopedia of Educational Philosophy and Theory*

Chapter No: 128-1

<table>
<thead>
<tr>
<th>Query Refs.</th>
<th>Details Required</th>
<th>Author’s response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU1</td>
<td>If you are using material from other works please make sure that you have obtained the necessary permission from the copyright holders and that references to the original publications are included.</td>
<td></td>
</tr>
<tr>
<td>AU2</td>
<td>Please change sentences with first person usage to third person.</td>
<td></td>
</tr>
<tr>
<td>AU3</td>
<td>Please provide “Synonyms,” if applicable.</td>
<td></td>
</tr>
<tr>
<td>AU4</td>
<td>Einav and Levin (2014) has been changed to Einav and Levin (2013) as per the reference list. Please check if okay.</td>
<td></td>
</tr>
<tr>
<td>AU5</td>
<td>Kitchin (2013) has been changed to Kitchin (2014) as per the reference list. Please check if okay.</td>
<td></td>
</tr>
<tr>
<td>AU6</td>
<td>Please check if edit to sentence starting “This use of…” is okay.</td>
<td></td>
</tr>
<tr>
<td>AU7</td>
<td>Please provide details for Driscoll and Walker (2014) in the reference list.</td>
<td></td>
</tr>
<tr>
<td>AU8</td>
<td>Please provide page range for Housley et al. (2014).</td>
<td></td>
</tr>
</tbody>
</table>