Chapter 4
Big Data and Doctoral Research: Opportunities, Challenges, and Cautions

Richard C. Berry
Intersect Australia, Australia

Lucy Johnston
University of Newcastle, Australia

ABSTRACT
This chapter explores opportunities and challenges that are presented to doctoral candidates (and indeed all researchers) through access to big data. The authors consider what big data is and what it is not, and how working with big data differs from traditional research design and analysis. They provide examples of the opportunities that big data offers in terms of the combination of diverse data sets, sources, and types and how it can provide new perspectives on inter-disciplinary challenges. They also highlight some of the challenges for the use of big data, both for the individual researcher and for institutions. The authors advocate for the need to embrace these challenges but without foregoing data integrity and the expert use and interpretation of data.

INTRODUCTION
Consider this thought experiment: How you would attempt to identify those people who do not default on credit-card payments. With concerns about personal debt and loan sharks, this ability would benefit not just credit-card companies but also broader society. Social scientists might consider retrospective comparison of

DOI: 10.4018/978-1-5225-7065-3.ch004
groups of individuals who have and have not defaulted on payments in order to identify differences between them. They may also conduct prospective studies that follow the activities of individuals with credit cards to see who does and does not default on payments. In the former, what variables would they compare between the defaulters and non-defaulters? In the latter, what aspects of the individuals’ lives would be tracked and how many people would have to be followed, and for how long, to get a large enough sample of defaulters and non-defaulters to give the study sufficient power? Whichever of these approaches, or any of the myriad of other possible approaches is taken, it is likely that it would be a long time, if ever, before such researchers identified that the purchase of anti-scuff furniture pads is a strong predictor of not defaulting. Identification of this factor is an example of the power of Big Data (Shaw, 2014).

The ease and reduction in cost of storing large quantities of data, and the huge improvements in the accessibility and discoverability of online data sets has led to a previously inconceivable increase in the availability of data which has had, and will continue to have, profound impacts on research, including doctoral research. Doctoral candidates are no longer restricted to using data that they are able to collect, store, and process in the limited timeframe and resources of their doctoral candidature. Instead they will have access to national and international datasets and networks of researchers that will facilitate not only access to more data, but also comparisons across time and location, and have opportunities for developing previously unconceived of research collaborations. Data from doctoral candidates’ research will also add to the global collective, where it can be made available to others such that the impact and influence of that data may be wider than that which might be achieved by the doctoral candidate working in isolation. For example, the interRAI project (www.interrai.org) looking at the lifestyle and health of the elderly has a standard 250 question survey administered to all participants in over 30 participating countries, including Australia and New Zealand. Doctoral candidates working on this project accordingly have access to a much larger data set than would ever have been possible prior to the developments in data storage, analysis, and in the ease of international communication and data transfer.

This increased access to data has had, and will continue to have, a positive impact on doctoral research, its advancement of knowledge, and impact on society. At the same time, however, this increased data access and storage will lead to differences in the doctoral research experience. The traditional doctoral thesis was an isolated piece of research with data collection, collation, and analysis often done by the candidate under supervision. The candidate typically took responsibility for the research design, the nature of the data collection, and for data integrity. Using Big Data may constrain the candidate’s control over these aspects, perhaps resulting in more emphasis on collaboration (including virtual collaboration), and inter-disciplinarity.
Discovering Patterns using Process Mining
www.igi-global.com/article/discovering-patterns-using-process-mining/163101?camid=4v1a