The rise of Big Data: what does it mean for education, technology, and media research?

Rebecca Eynon [a]

[a] Oxford University , Oxford , UK
Published online: 25 Feb 2013.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

EDITORIAL

The rise of Big Data: what does it mean for education, technology, and media research?

The interest in Big Data is growing exponentially. Research calls, commercial insights, and government initiatives all seem to be focused on exploiting the potential of technology to capture and analyse massive amounts of data in increasingly powerful ways. Big Data, that is, data that are too big for standard database software to process, or the more future-proof, 'capacity to search, aggregate, and cross-reference large data sets' (boyd and Crawford 2012, 663), is everywhere. For some, Big Data represents a paradigm shift in the ways that we understand and study our world, and at the very least it is seen as a way to better utilise and creatively analyse fine grained data for public and private benefit.

In some ways, this is not a new phenomenon. Those working in the commercial sector have been collecting and combing large data sets to improve segmentation of goods to customers and better understand their market for many years (Manyika et al. 2011). Nor is it particularly big news to those working in certain fields in the natural sciences. Yet, in recent years, a far wider range of stakeholders have become more involved and more excited about the potential of Big Data.

In the field of education, Big Data is still a relatively niche topic, but it is clearly beginning to grow. The areas of Educational Data Mining and Learning Analytics are both developing a number of identifiers characteristic of an established field of study (for a very useful overview see Romero and Ventura 2010; Ferguson 2012); and governments are beginning to produce reports on the potential of Big Data for education (see, for example, the 2012 consultation report by the US Office of Educational Technology, Department of Education).

However, perhaps unsurprisingly, the pervading discourse around Big Data and education is increasingly one of efficiency and cost-effectiveness, both in terms of the use of the data to improve education 'delivery' and as a means of carrying out research in our field. While there is some merit in these claims, it is important that discussions of using data to enhance efficiency, increase transparency, support competiveness, and as a tool to evaluate performance (of schools and teachers) are tempered with considered academic debate. In times of austerity, commercial discourses become more significant in our debates about education. Thus, in some quarters Big Data is increasingly promoted as a form of 'technical fix' for education research and practice (Robins and Webster 1989), rather than a focus on using these same tools to empower, support, and facilitate practice and critical research.

This is not to say Big Data is not important, or that this is not an exciting development for research and practice in the Learning, Media, and Technology field. Yet, as a community we need to shape the agenda rather than simply respond to the one offered by others (see also Drachsler and Wolfgang 2012). To assist with this debate, three areas meriting attention are highlighted below.

The first sets of challenges are to do with ethics. Big Data represents a number of ethical considerations, particularly around privacy, informed consent, and protection of harm, and raises wider questions of what kinds of data should be combined and analysed, and the purposes to which this should be put.

All of these considerations are vital, and many of these debates have been documented elsewhere. However, the question of how the data are used has been less often debated in the educational sphere. For example, it is already possible, using learner analytics, and educational data mining techniques, to identify who at registration is likely to drop out of an online higher education course. Using those same techniques, it is also feasible to design systems that offer students a menu of learning options, similar to the way we are offered books and other products by online shops, 'students who took the courses you took also took (…)' or 'students who did well in these courses also did well in (…)' Using Big Data, we could better identify failing schools or failing teachers (and some of the reasons why), and we could potentially track almost everything a student did while in school.

Whilst capturing and analysing all of these kinds of data could have positive effects for research and practice, we need to think carefully about the social implications of this kind of use of the data. What happens to students who are most likely to drop out? Do we tell them, support them (which has economic implications), or let them sign up and take their money, knowing they will probably fail? What happens to serendipity in a system where all educational choices are based on recommender systems? What kinds of learning can a student truly keep 'private'? Does the potentially highly public and trackable nature of learning have impacts for the learning process? There are any number of questions of this kind, which need to be asked and critically considered every time data is analysed and used. Data, particularly Big Data, have a kind of kudos that needs to be treated with care, as the values that are designed in to the analysis process are not always properly considered or made explicit.

A second set of related challenges are about ensuring an understanding of the kinds of research that can or cannot be carried out using Big Data. The availability of Big Data limits the kinds of questions we can ask, as we can only study data we can collect or already have. We do not have data for everything, thus the availability of Big Data shapes what people research and the questions that they ask. In addition to issues of availability, we also typically only have access to certain kinds of data. Big Data is essentially capturing and examining patterns, and typically it tells us more about what people actually do than about

what they say they do (as is common in surveys and other more well-established social science techniques). However, whilst being able to measure what people actually do is of great value, this is not sufficient for all kinds of social science research. We also need to understand the meanings of that behaviour which cannot be inferred simply from tracking specific patterns.

For example, in a recent research project, I was involved in we spoke to young people who said (in surveys) that they had stopped using the internet. In interviews, they said the same thing, but with greater probing told us that they did actually use it occasionally (for *Facebook* and job search most typically). Perhaps, the use of Big Data would have told us that they did, in fact, use the internet – but the 'stopping' of the internet in their terms meant something important, highlighting how little meaning and/or cause for frustration the internet had in their lives (Eynon and Geniets 2012). Likewise, when studying relationships, some of the connections between people can be measured (e.g. address book contacts, number of emails sent). However, this does not equate to the value or meaning people place on those connections (boyd and Crawford 2012). Meaning is crucial in our understanding of the world and this must not be forgotten when conducting Big Data studies. Instead, we need to think about how to ensure a more holistic approach to research in this field.

The final set of challenges are those around issues of inequality, and how Big Data may both reinforce and perhaps even exacerbate existing social and educational inequalities in a number of ways. First are issues around the question of whose data traces will be analysed using Big Data, and in simple terms it is likely that only those who are better off will be represented in such research, as these are the people who will be online more. For example, a lot of work in Big Data focuses on *Twitter*, the blogosphere, and search engine queries. All of these activities are not undertaken equally by the whole population. Second are issues around the question of which researchers have access to these data sets, which are often owned by commercial companies (boyd and Crawford 2012). Similarly, access and use of open data is unlikely to be equally available to everyone due to existing structural inequalities.

In addition to these questions there are a number of practical issues. These include: implications for the training of future academics in our field (how skilled must we all be at this kind of analysis, and is it sufficient to learn to work with others in multi-disciplinary teams?); institutional and cross-institutional infrastructure to support data storage and processing capacity; agreements and incentives for sharing data across a range of different actors; ethical guidelines and protocols – and so on. It is important that issues do not become the core focus of our activities and debate, but we cannot afford to ignore them.

Big Data offers some really interesting opportunities for research and practice. It could help us to better understand a range of issues at the heart of learning and technology research, and as a community we are well placed to make

use of this data in creative ways. However, we must not get seduced by Big Data. It is easy for scale and processing power to look impressive. Indeed, the conclusions from Big Data studies are likely to offer some neat answers at first glance, particularly in comparison to the messiness of more established social science approaches. But we need all of these approaches to truly understand phenomena. The debate needs to be about more than overcoming 'barriers' to Big Data (personnel, equipment, cultural mind sets, etc.), so much as about ensuring that we use these techniques to empower researchers, practitioners, and other stakeholders who are working in the field.

References

boyd, D., and K. Crawford. 2012. "Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon, Information." *Communication & Society* 15 (5): 662–679.

Drachsler, H., and G. Wolfgang. 2012. "Confidence in Learning Analytics." Learning Analytics and Knowledge 2012 Conference, Vancouver, Canada, April 29–May 2.

Eynon, R., and A. Geniets. 2012. *On the Periphery? Understanding Low and Discontinued Internet Use Amongst Young People in Britain*. Report for the Nominet Trust, Oxford, UK.

Ferguson, R. 2012. *The State Of Learning Analytics in 2012: A Review and Future Challenges.* Technical Report KMI-12-01, Knowledge Media Institute, The Open University, UK. Accessed January 21, 2013. http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey: McKinsey Global Institute.

Office of Educational Technology. 2012. *Expanding Evidence Approaches for Learning in a Digital World*. Washington, DC: US Department of Education, Office of Educational Technology. Accessed January 21, 2013. http://www.ed.gov/edblogs/technology/files/2012/12/Expanding_Evidence_Approaches_DRAFT.pdf

Robins, K., and F. Webster. 1989. *The Technical Fix: Education, Computers and Industry*. New York: Macmillan.

Romero, C., and S. Ventura. 2010. "Educational Data Mining: A Review of the State of the Art." *IEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* 40 (6): 601–625.

Rebecca Eynon
Oxford University Oxford, UK