Data entry: towards the critical study of digital data and education

Neil Selwyn
Monash University, Melbourne

ABSTRACT: The generation and processing of data through digital technologies is an integral element of contemporary society, as reflected in recent debates over online data privacy, 'Big Data' and the rise of data mining and analytics in business, science and government. This paper outlines the significance of digital data within education, arguing for increased interest in the topic from educational researchers. Building on themes from the emerging sub-field of 'digital sociology', the paper outlines a number of ways in which digital data in education could be questioned along social lines. These include issues of data inequalities, the role of data in managerialist modes of organisation and control, the rise of so-called 'dataveillance', and the reductionist nature of data-based representation. The paper concludes with a set of suggestions for future research and discussion, thus outlining the beginnings of a framework for the future critical study of digital data and education.

Keywords: digital data, education, analytics, measurement

Data entry: towards the critical study of digital data and education

INTRODUCTION

The prominence of data as a social, political and cultural form has risen significantly in recent years. Of course, the process of collecting measurements, observations and statistics together for reference and/or analysis has taken place for centuries. Yet the past twenty years or so have seen the increased recording, storage, manipulation and distribution of data in *digital* form (usually through computers). In this sense, digital forms of data are now being generated and processed on an unprecedented scale. This shift is often described in terms of 'three Vs' of volume, velocity and variety – i.e. increases in the amount of data that is now being produced; the speed in which this data can be produced and processed; and the range of data types and sources that now exist (Laney 2001). Yet digital data is also distinct from pre-digital forms by being exhaustive in scope, highly detailed and flexible in the ways that it can be combined (Kitchin 2014). Indeed, the constant circulation and reconstitution of digital data that now takes place has prompted talk of a 'data deluge' (The Economist 2010) and 'data flood' (Palmer 2012). For better and for worse, then, the generation and processing of digital data is an integral element of many aspects of contemporary society. As such, this is a topic that clearly merits the attention of educational researchers and other social scientists with an interest in education.

THE RISE OF DIGITAL DATA AS A SOCIETAL AND EDUCATIONAL PHENONEMON

It is first necessary to be clear of what 'digital data' is [1]. We can begin by noting that digital data originates from a variety of different sources and takes a number of different forms. On one hand, many forms of digital data are generated deliberately – such as the administrative measurement of people's interactions with government agencies and commercial interests, as well as more covert forms of institutional monitoring, surveillance and evaluation of individuals. Alternately, large volumes of data are generated and archived automatically through the 'routine operations' of digital devices and systems (Savage and Burrows 2007). Vast quantities of data are also volunteered by individuals in the course of their use of digital technologies. This includes data arising from the routine curation of personal 'profiles' on social media as well as various other forms of 'user-generated content'. As Bruno Latour (2007, n.p.) notes, "the slightest move in the virtual landscape has to be paid for in lines of code".

All these different forms of data originate in a 'raw' state of machine-readable alphanumeric symbols. Yet once raw data begins to be processed it then takes the form of more socially meaningful 'data entities' – i.e. representations, models and calculations relating to 'real world things' such as people, places or products. This processing occurs through a variety of computer-based and often automated 'data work' – i.e. the archiving and tagging, linking and connecting, 'harvesting' and

'mining' of data through various computational processes. This allows institutions (and more specifically individuals working on behalf of institutions) to engage in computational work that aims to predict, model and distil data for human judgement. In particular, the recent turn towards 'Big Data' refers to these processes taking place on a mass, aggregated scale. Now, the ever-increasing computational capacity of digital systems offers opportunities to work with huge amounts of data and discover patterns that would otherwise be imperceptible (Boyd and Crawford 2012).

The growth of digital data production and processing has been accompanied by a set of widely expressed public hopes *and* fears. For example, recent revelations regarding the activities of the US government 'National Security Agency' gave renewed impetus to long-standing public concerns over data-related surveillance, privacy and monitoring. Additionally, as terms such as 'deluge' and 'flood' imply, are fears that populations are simply being overwhelmed by excessive quantities of data. Counter to such pessimism, however, are prevailing claims for the societal benefits of data processing. For instance, it is argued that expanded access to data allows institutions and individuals to operate more efficiently, effectively and equitably. It is also argued that increased data access can democratise decision-making processes, make institutions more 'transparent' and elite actors more 'accountable' for their actions. Other benefits are also seen to stem from the connections and linkages that can be made between previously disparate and disconnected sources of information – what has been termed 'combinatorial innovation' (Yoo *et al.* 2012). Many of these perceived advantages reflect an underlying belief that digital data renders social processes and social relations more knowable and, it follows, more controllable. As Nick Couldry (2013, n.p.) concludes, this increasing trust in the power of data and digital technology "has already rationalized a state of affairs where a network of data-gathering and data-amalgamating institutions has, or aspires to have, everything".

For better and for worse, then, digital data is now an integral feature of government, scientific work and commercial activity. Yet, while less discussed than the high-profile areas of 'Big Data Science' and 'Business Intelligence', it is also worth acknowledging the ways in which education has been subject to a similar digitally-driven 'datafication'. Indeed, schools, colleges, universities and other educational contexts now function increasingly along 'data driven' lines. Within even the smallest of schools, for example, masses of digitized data are being generated, collected and collated on a daily basis. These data range from the often *ad hoc* 'in-house' monitoring of students and teachers to the systematic 'public' collection of data at local, state and federal levels. Vast amounts of 'naturally occurring' data are generated from the daily use of 'virtual learning environments' and other learning technologies that log information during the course of their operation. All these data are then processed for a variety of purposes – including internal school administration, target setting, performance management and student tracking.

Most significant, perhaps, are the pervasive forms of data work that now exist across regional and national educational systems – from the curation of national student databases, the processing of examination results, school performance 'league tables' and the collated use of school inspection reports. The capacity to monitor targets and create league table positionings is now a key aspect of national and international policymaking, leaving digital data in education a "relentless and inescapable" feature of contemporary education governance (Ozga 2009, p.154). The political significance

of educational data is evident, for example, in the provenance afforded to global 'performance indicators' such as PISA and TIMSS. The normalisation of digital data work within education is also apparent in the growing acceptance of 'learning analytics' throughout compulsory and post-compulsory education – i.e. "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens *et al.* 2011, p.4). There is also growing interest in the field of 'educational data mining', which promises to make use of "data to enhance efficiency, increase transparency, support competiveness, and as a tool to evaluate performance" (Eynon 2013, p.237). All told, the generation, accumulation, processing and analysis of digital data is now being touted as a potential panacea for many current educational challenges and problems.

TOWARDS THE CRITICAL STUDY OF EDUCATION AND DATA?

It could be argued, therefore, that contemporary education cannot be understood fully without paying proper attention to the accumulation and flow of data. In particular, the notion of a contemporary educational landscape infused with *digital* data raises the need for detailed inquiry and critique. While educational research has been generally slow to respond to the rising significance of data, one exception has been work within the area of policy sociology in addressing the changing nature of educational governance. This work has charted the rising prominence of data within various education systems, allied to the new managerial 'governance turn' of the past twenty-five years. Jenny Ozga (2009), amongst others, has shown how the use of data has been particularly notable in the growing use of goals, targets, benchmarking, measurement, performance indicators and monitoring within the English education system. Data-related technologies of governance have been also noted in the Australian context, with Bob Lingard exploring the rise of he terms 'policy by numbers' (Lingard *et al.* 2012). Similarly, Grek and Ozga's (2010) comparative European research has shown how data now plays a key role in attempts to 'harmonize' the complex European education landscape, with data systems being used to "construct policy problems and frame policy solutions beyond and across the national scale" (Ozga 2012, p.440). Similarly, Gorur's (2014) research into the construction of OECD's Programme for International Student Assessment (PISA) highlights the "instrumentalism and performativity" of the data-driven technologies of international comparisons. In all, these studies provide a good account of the ways in which data production, data management and the associated state of 'constant comparison' now underpin how education systems are now governed and controlled (Ozga 2009, p.150).

Another policy-related body of research on education and data is the growing number of studies influenced by an Actor Network approach. In focusing on the assemblages of human and non-human actants that constitute any educational environment, Actor Network studies have invariably included data processes and practices as key elements of these assemblages. Gorur and Koyama's (2013) study of data-driven 'like-school' comparisons in US and Australia, for example, outlined the growing use of numeric calculations to produce 'simplified technical accounts' (p.633) of educational problems and their 'seductively simplistic' solutions (p.645) – pointing in particular to the ways in which a calculative mindset has begun to form a subtle

'technicising' infrastructure underpinning the ways in which education is understood and organised. Similarly, Koyama and Menken's (2013) study of the counting and tracking of immigrant students in the US showed how these students were being positioned as a cause of school and system-wide 'failure' through the 'numerical manipulation' (p.83) of test scores, district-level data and other reporting mechanisms. Here, data was seen to narrow the efforts and attention of education organisations to be "myopically centred on generating, gathering and reporting data – with detrimental effects" (p.83). Koyama's (2011) study of for-profit 'supplemental educational services' in the US also highlighted how calculation (notably of test data) played a key role in sustaining – and legitimizing - the network of schools, private companies and district officials that formed around this new aspect of school provision. As Fenwick and Edwards (2011, p.722) outline in more detail, Actor Network studies such as these do much to identify and problematize the complex connections and connotations of any piece of seemingly 'static' piece of data:

> "For instance, if we think of a national standardized test score, at any given moment it balances a number of materials, forces, and values: a policy put in place to standardize assessments, the negotiations and consultations on the nature and form of those assessments, a package of test forms delivered to a school, a seated child writing the test, a worried parent, a test-item database, a list of grades, a school inspector examining the grades, a league table calculating international test comparisons, and so forth. Continuous effort is required to hold an assessment-network together, to bolster the breakages and counter the subterfuges".

Beyond these existing bodies of research, however, there is much more that educational research can do to respond to the increasing prominence of digital data within contemporary education. In particular, more work is needed that develops a socially sophisticated and robust understanding of the 'digital' characteristics and qualities of educational data work and how various forms of digital data are set to work within educational contexts. This clearly calls for a technical, as well as sociological, sensibility. As such, there is much that might be appropriated from the emerging interest in mainstream sociology in digital practices and digital cultures. Indeed, readers of general sociology journals and/or attendees at sociology conferences might have noticed the burgeoning presence of researchers and writers from across Europe, North America and Australia aligning themselves around the label of 'digital sociology'. Key writers in this vein include Susan Halford and colleagues at Southampton University, Evelyn Ruppert at the Open University, and Kate Crawford at MIT. Academics associated with the University of York such as Mike Savage, David Beer, Andrew Webster and Roger Burrows have also worked to raise the attention of British sociology to the implications of digital data, as have Deborah Lupton and Theresa Sauter in Australia. Cognate work has taken place in human geography (for example, Rob Kitchin and Mark Graham) and the humanities-orientated field of 'software studies' (notably Lev Manovich and Matthew Fuller). These individuals – and others like them – are doing much to develop technically-aware understandings of digital data within the general discipline of sociology. It therefore follows that there is much that might be appropriated and drawn upon in future educational research.

As with most sociological studies of technology, these researchers and writers are all striving to open up the 'black box' of digital data. This contrasts starkly with popular understandings of 'data' to be broadly neutral, objective and therefore non-problematic in nature. Instead, the digital sociology approach tends to start from the contention that data is political in nature – loaded with values, interests and assumptions that shape and limit what is done with it and by whom. Yet this approach is careful to acknowledge that data is profoundly shap*ing* of, as well as shap*ed* by, social interests. Indeed, it is argued that computer code, data and other 'digital informational flows' are increasingly defining as well as describing social life. As Beer and Burrows (2013, p.63) contend:

> "the 'stuff' that makes up the social fabric has changed … social associations and interactions are now not only mediated by software and code but they are also becoming increasingly constituted by it".

In this sense, much of the sociological significance of digital data is seen to relate to its association with meaning-making. Of course, most forms of data are concerned inherently with attempts to make sense of the social world and understand the 'way things are'. Thus as Nick Couldry (2013, n.p.) observes, recent popular debates around computerized data are rooted in wider debates about defining what is known and what is knowable in contemporary society – i.e. "what counts as social knowledge". As such, on-going debates about digital data relate implicitly back to broader struggles of how it is possible to 'establish truth' in contemporary society (Graham and Shelton 2013). Data processes that might seem mundane and procedural are often significant and highly powerful social practices (e.g. processes of observing, measuring, describing, categorising, classifying, sorting, ordering and ranking). An obvious point to make from a sociological perspective, therefore, is that these processes of meaning-making are never wholly neutral, objective and 'automated', but are fraught with problems and compromises, biases and omissions. As we shall go onto discuss, some of the key concerns over digital data are therefore those of representation (with finite sets of characteristics being decided to 'count' as a particular entity) and reductionism (with artificially neat boundaries and categories being drawn around data). As Halford *et al*. (2013, p.180) conclude, "in short, the processes involved in naming, structuring and processing data … are profoundly social with tremendous sociological implications".

The inference here, then, is that digital data should be seen as playing a key part in defining as well as merely describing 'the social'. As Beer and Burrows (2013, p.64) contends, "data is recombinant and recursive, it shapes as well as merely captures culture". This suggests paying particular attention to the ways in which digital data 'circulates back recursively' into everyday life and everyday cultural forms (*ibid*. p.67). A key concept in this respect is gaining better understandings of the 'social life' of digital data – i.e. the continual re-use and re-constitution of data into different and new forms. In this sense, digital data should not be seen in simple terms as something that is collected and then used in a single discrete action. Instead, diverse sets of raw data are being continually combined and recombined, with different data entities produced from varying iterations and calculations. In short, any form of digital data is an evolving entity that the original sources often have little or no control over. As Andrew Webster (2013, p.230) concludes, "data itself can take on its own life …

these data then travel, are transformed and are transcribed into novel 'derivative' forms".

These are all significant concerns that are now only beginning to be articulated and explored. Indeed, it would be misleading to suggest that a coherent 'sociology of data' already exists that has fully theorised and made sense of such questions. At present, this is a nascent but clearly important area of sociological thinking. Moreover, this is an area of sociology that educational research could – and should – be playing a part in shaping. As the remainder of this paper will now go on to argue, there is much in the topic of digital data that maps onto the core concerns of educational research. All of the issues just outlined highlight the need for better understandings of how uses of digital data are implicated in the shaping of what people can and cannot do, in the shaping of opportunities and – in short - in the operation of power. Therefore, there is clearly a need for further sustained work within educational research that begin to explore what it means to live and work within the data 'deluged' conditions of educational settings described earlier. As Beer and Burrows (2013, p.75) conclude:

> "[social scientists] need to find ways of getting to grips with the informational infrastructures and how these mesh into those established concrete structures and geographical social patterings … the need to understand the construction of code and its operation from a sociological perspective becomes fundamental".

AREAS OF EDUCATIONAL CONCERN EMERGING FROM THE SOCIOLOGY OF DATA

These calls are now beginning to be responded to across many areas of the social sciences. In this spirit, there are a number of concerns with digital data that have clear relevance to contemporary education and educational research:

i) Digital data and the reproduction of inequalities and social relations

First is a concern with how digital data is implicated in the reproduction of existing social inequalities, as well as implicated potentially with the generation of new forms of inequality. A central concern of much of what has been discussed so far in this paper relates to social power and control being reinforced, or perhaps reconstituted, through data-driven processes. It has been suggested, for example, that the increased presence of digital data systems throughout society make power ever more 'invisible' and 'taken-for-granted' (Lupton 2013). Moreover, is the potential for 'new' intensifications of inequalities of power and control arising from the generation, processing and circulation of digital data. Like many topics of sociological significance, digital data need to be problematized in terms of power, control, domination and inequality.

At a basic level, then, we need to acknowledge the unequal agency that individuals and social groups have when engaging with digital data. Put crudely, a distinction can be made between those who merely have data 'done to them', as opposed to those

who have the ability to 'do data'. In this sense, Lev Manovich (2011) has pointed to a new hierarchy of 'data classes' associated with the increased use of digital data in society. This spans from the majority of individuals who simply create data for others to process (and are largely unconscious of doing so); those who create data but are often conscious of doing so; those who have the means to collect data; and finally those who have the expertise to analyse data. Clearly, these different groups are ordered along lines of technical and statistical expertise – what Manovich (2011) has described as a new 'data analysis divide' between data experts and those without computer science training. Yet this hierarchy also maps onto existing power differentials and unequal social relations. As Ruppert (2013) notes, it is notable how the dominant 'datascapes' of contemporary society tend to be tied closely with dominant 'theories of social order'.

These issues are increasingly apparent in the permeation of digital data throughout educational contexts. On one hand, cadres of data-analysts and technocrats are now employed in universities, colleges and (to a lesser extent) schools to deal with the processing and 'doing' of digital data. Data processing is therefore often experienced as institutionally-driven and 'top-down' in nature. Conversely, many students and teachers remain largely unconscious of the extent and implications of their daily production of digital data traces and trails. This clearly raises the question of who is able to benefit from data work in educational contexts. Militello's (2013) study of how assessment data were used within US schools highlighted a stark disparity between classroom teachers (who tended to respond to data in a self-regulatory sense, seeing it as indicating changes required of their own practices), and school principals (who tended to see data primarily as indicating changes required to the work of others). In both sets of responses, assessment data tended to be perceived as an end (rather than means) of educational work. Clearly more work is needed along these lines – exploring how digital data systems are encountered and experienced by different groups of actors within educational contexts.

ii) Digital data and the intensification of managerialism within education

These concerns over power, control and performativity are allied with the role of digital data in reinforcing and intensifying the culture of managerialism within education. Clearly, digital data is now a core element of managerialist techniques of accountability, auditing, evidence-based management, 'evidence based' practice, effectiveness and so on. In their heighted use of digital data, schools, colleges and universities are therefore joining the ranks of other contemporary institutions such as hospital, businesses and prisons in infusing themselves with 'organisationally focused' uses of data and information. Borrowing Thrift's (2005) notion of 'knowing capitalism', it could be argued that digital data are supporting a new form of knowing managerialism within educational settings. This describes the role of managerial interests in gathering data and information in an on-going attempt to make sense of the everyday – i.e. to "consider its own practices on a continuous basis ... to use its fear of uncertainty as a resource ... to circulate new ideas of the world as if they were its own ... to ... make business out of, thinking the everyday" (Thrift 2005, p.1).

Of course, the continual collection and analysis of data has always been a central tenet of managerialism. As Bowker and Leigh Star (1999) detailed, data has long sustained

a managerialist culture of 'sorting thing out' within large institutions – as seen, for example, in the organisational classification of individuals as users, clients, patients and/or consumers. Yet there is a sense that digital data have extended and intensified these processes. On one hand, digital data have introduced the sense that organisations such as schools and entire sectors of society such as education can be seen as 'computational' projects (Kling 1991). Here, the 'modelling' of education through digital data is seen to engender a sense of algorithmically-driven 'systems thinking' – where complex (and unsolvable) social problems associated with education can be seen as complex (but solvable) statistical problems. Thus digital data are accompanied by a heightened sense of 'solutionism'. This leads to a recursive state where data analysis begins to produce educational settings, as much as education settings producing data. This state of interdependence reflects Rob Kitchin's notion of contemporary society as 'code/space' – i.e. where real-world spaces and software code "become mutually constituted, that is, produced through one another" (Kitchin and Dodge 2011, p.16).

A few recent studies have explored the ways in which digital data are entwined with the reinforcement of managerial conditions in education. For instance, a study into the 'evidence-based' closure of a US school serving predominantly African-American community pointed to the oppressive and unjust consequences of data-driven decision making by school authorities who think that they are "merely enacting technical-administrative behaviours" (Khalifa *et al.* 2014, p.148). Conversely, the role of data in the politics of university organisations has also been well noted. Ayers (2014), for example, shows how the tensions and struggles of higher education governance are played out through essentially abstracted data-based struggles such as 'budget updates' and other intra-organizational 'data wars'. As Browne and Rayner (2013, p.7) conclude:

> "data are now being generated and then recorded at every level of educational activity in the university, and framed as a purposeful re-production of data-mining, spin and marketing, all reflecting the further expansion of a managerialist, evidence-informed industry represented in a burgeoning technicist mantra of educational authentication and workforce accountability".

iii) Dataveillance

Digital data can be used against those working in educational contexts in a variety of ways. Alongside these managerialist applications of data is the rise of so-called 'dataveillance'. Of course, the role of digital technologies as direct tools of surveillance within educational institutions has been well documented, with schools, colleges and universities now replete with surveillance technologies from CCTV and RFID tracking through to the monitoring of internet use (see Taylor 2013). In contrast, dataveillance constitutes what Torin Monahan (2010, p.86) terms the 'surveillance of abstract data' – a seemingly 'less intrusive and less threatening' form of monitoring within educational institutions. The rise of data surveillance on both a personal and mass scale has been well documented over the past thirty years, describing the process of monitoring the 'data traces' that an individual leaves when using digital media (Clark 1988). Often this monitoring is conducted surreptitiously (as in the case of webpage 'cookies'), but also often takes place on an unwittingly

permissive basis. For example, Albrechtslund (2008) describes the data that users volunteer through the updating of social media, digital calendars, user profiles and the like as a mass 'participatory surveillance'. Indeed, the essentially invisible and continuous nature of this data collection has led to it being described as a form of pervasive 'silent control' (Orito 2011).

Crucially, this monitoring, mining and processing of data supports a range of data-profiling processes. Indeed, the data processing arising from dataveillance allows for the identification, classification and representation of social entities (be they people, places or events) in the form of automated data profiles - sometimes described as 'data doubles' or 'data shadows'. As Tække (2011, p.446) observes, "the human actor of surveillance is replaced with a computer system that constructs the public observable self independent of our present self-presentation". Crucially, this knowledge building is then used to support 'predictive' profiling, where the future behaviours of an individual can be calculated and then acted against pre-emptively. Gandy (2012) refers to this as 'statistical surveillance', with computer analysis of statistical data providing institutions with 'actionable intelligence' to underpin decision making and choices. Thus data are used for a number of 'determinations' – identifying who an individual is, classifying what they are, and evaluating what they might be. This predictive determination can lead to a variety of 'statistical discrimination', where individuals are reclassified in terms of their associations and linkages with others, and then including/excluding on the basis of the attributes of the groups and data 'segments' that they belong to.

Such dataveillance practices are prevalent within educational contexts. As Rosenzweig (2012) notes, continual dataveillance of digital technology use is an accepted 'condition of employment' for teachers. Conversely, dataveillance is now embedded into most technology-based forms of teaching and learning encountered by students. As Emmaline Taylor (2013, p.9) has noted, "by embedding surveillance into pedagogical apparatus, young people are being habituated to unprecedented levels of scrutiny and control". These latter forms of dataveillance have attracted some attention from social researchers. For example, Land and Bayne (2005) discussed the 'student tracking' capabilities of virtual learning environments, noting these systems' collection of 'sophisticated' data trails from students and tutors under the aegis of its 'pedagogical functioning'. While this state of dataveillance is constructed as "useful ways of evaluating course effectiveness through helping us to understand student usage of the online facility" (p.165), Land and Bayne argue that such regimes of tracking and surveillance impact on the individuality of learners, fostering specific subjectivities and modes of self-governed behaviours. Similarly, Dan Knox (2010) has pointed to how the automated surveillance and heighted visibility implicit in online learning environments in higher education leads to a form of 'coded suspicion' between academic staff, administrators and students. This can lead in turn, Knox contends, to a corrosion of organizational trust which impacts negatively areas such as work effort, quality of dialogue, academic achievement and intellectual risk-taking.

iv) Digital data and the reductive nature of 'what counts' as 'education'

Underlying these latter points is the question of what is being lost in the educational turn towards digital data. This concern is writ large throughout the general

sociological writing on digital data, with researchers keen to explore the reductionism associated with different forms of data. Any data entity is built around the assumption that most (if not all) characteristics of the social entity that it purports to represent can be measured and represented in a discrete and decontextualized manner. Thus any attempt to label, structure and assign values to data is a finite and limited process of representation and interpretation. For example, while digital systems rely on the assumption that "a human being is considered to be a data set" (Orito 2011, p.9), the process of compiling a representative data profile about any individual is clearly fraught with difficulty. All processes of data collection and recording involve "categorical thinking based on the binary either/or logic, dominates, which puts individuals into categories and, in the process, obscures any ambiguities" (Parton 2008, p. 263). The effective use of digital data therefore relies on making a number of assumptions that do not necessarily reflect the complexities of social life. As Scott Lash (2002) concludes, the information gathered for data systems often based upon concerns of operationality rather than nuance of social meaning.

Digital data (and its analysis) therefore needs to be seen as a contestable process - "often unreliable, prone to outages and losses" (Boyd and Crawford 2012, p.668). It is necessary to consider what tends to be underrepresented or excluded altogether in digital data sets. Of key concern to sociologists, is the tendency of digital data to remove 'the social' from acts of knowing – i.e. "elements that connect with how individuals, with recognisable sets of human aims and capabilities, make sense of what they do" (Couldry 2013, n.p). The recording of social 'facts' into digital data therefore implies that some qualities and characteristics will be made better known than others. For example, as Ruppert (2012) notes, the core sociological constructs of race, social class, gender, sexuality and so on, do not translate easily into data categories, despite their constant use within data collection and analysis. Often digital data can be said to support little more than 'surface' understandings of social entities (Savage 2009). Indeed, Manovich (2011) highlights the differences between 'deep data' about a few cases and 'surface data' about a large number of cases. Much of the depth that is lacking from digital data could be argued to include issues of historical context and connections with past events, individualist and humanist accounts of the social, and an underpinning sense of moral knowledge (see Barnes 2013, Ruppert 2013).

Questions therefore need to asked regarding what reductions exist in how education is now 'known' through digital data. Clearly consideration needs to be given to the biases towards measuring what can be measured most easily. As Graham and Shelton (2013, p.258) reason, "because data are always constructed, collected, stored, and used under uneven and variegated social, economic, and technical contexts, some people, places, and processes will always be easier to enrol into such vast sociotechnical assemblages". In education this can be seen in the prominence of 'results' from formative assessment, inspection reports, 'direct' measures of attendance and so on. It is important to recognise the limitations of all such measures that are given prominence in contemporary education. For example, the crude, essentialised nature of indicators of 'student satisfaction' in higher education or the 'effectiveness' of secondary schools has been well discussed (Douglas *et al.* 2014, Gorard 2010). Similarly, Browne and Rayner's (2014) analysis of the 'key information set' data made available to university students in the UK highlights the highly partial 'story told' about institutions, and the partisan ways in which such data

are reused and represented. These data tended to omit a range of information about higher education relating to human individuality, the potential for growth, freedoms and trust associated with professionalism, greater social justice and "a humane and moral sense of the academic endeavour" (Browne and Rayner 2014, p.15).

MATTERS OF CONCERN FOR EDUCATIONAL RESEARCH

These brief examples highlight the significance that digital data is beginning to assume in contemporary education. With this in mind, the opportunity now exists for educational research to develop nuanced approaches to understanding, and then offering alternatives to, the dominant data conditions that are being established across educational contexts. First, this requires an appropriately critical approach towards researching digital data within educational contexts. This would be empirical work that strives to understand and account for the manner in which data is accumulated; to make visible the flow and circulation of data; and begin to understand the ways in which data are then integrated back into everyday education practices. There are a number of areas of research questioning that therefore demand sustained consideration – for example:

> *i) What data exist in educational contexts?* How are educational institutions and organisations gathering data? What areas of education do these data relate to (i.e. teaching and learning, organization and administration, leadership and management, change and innovation)? In what forms do these data exist and in what forms are these data accessible (e.g. open/closed access, raw/value-added)? Are these data created intentionally or 'naturally-occurring'? What are the quality, scope, inter-operability and compatibility of these data? What is assembled – included/excluded, present/absent, inside/outside of these data? How has this assembling of data varied and changed over time?

> *ii) What are the 'primary' uses of these data?* e.g. measurement, monitoring, formative or summative assessment? Where in education systems are these data being used, e.g. individual classrooms, departments, schools, colleges and universities, regions, states/provinces, or (inter)national? How are these data being used and by whom – i.e. data work by internal and external actors, data work for auditing and assessing, or decision-making and planning?

> *iii) What – if any - are the 'secondary' uses of these data?* For what purposes are these data being re-used and by whom? How are these data being used for prediction; analysing trends and patterns; modelling; distillation of data for human judgment? How do these data inform 'rules of thumb', informal models and implicit practices of understanding and making sense? What is the 'social life' of these data – i.e. how are these data being aggregated, segregated and reconstituted? What innovative data practices can be identified within educational settings?

> *iv) What are the consequences of these uses of data?* Are these data uses leading to improved outcomes, efficiencies, self-regulation and/or relationships? How are data uses related to alter social relations within educational contexts – i.e. in terms of relations of power and control,

conditions of performativity and/or surveillance? How do these consequences differ between students, teachers, administrators, educational leaders and managers? How are digital technologies supporting the connection, aggregation and use of data in ways not before possible?

v) What organizational cultures have formed around the use of data within educational settings, and with what outcomes? Where does data work mirror existing institutional structures and hierarchies? Where is data disrupting, changing or leading to new arrangements, relationships and understandings? Where is data leading to a refocusing in practice/understandings towards to measurable and 'visible'? What ethical, legal, managerial and organizational issues are shaping the use of data within educational settings?

vi) How might data-work be more efficiently and equitably arranged in educational contexts? How might authorities ensure 'beneficial' collective use of the total data that is available to those who currently have data 'done to' them, rather than having the availability 'to do' data? Conversely, how might data access and use be more democratically arranged across all elements of educational communities? In both these senses, what types and forms of data and data accessibility are desirable? How might quality data sets (in terms of scope, interoperability and compatibility) and sources be developed? What data tools and technologies are required?

The scope of these research questions also raises issues of methodological capacity. These are all questions that require educational researchers to develop the skills and attributes necessary to engage effectively and insightfully not only with the social uses of data, but with the data itself. In other words, this implies the development of analytic skills and competencies that might allow educational researchers to work *with* data as well as working *on* data. These skills include methodological approaches such as 'trace ethnography' (see Geiger and Ribes 2011, O'Keeffe 2014), data visualisation and other 'critical code' methodologies (see Fuller 2008). As George Siemens (2013, p.1390) notes, there is clearly a need for educational researchers to develop capacity in the areas of "programming skills, statistical knowledge, and familiarity with the data and the domain represented in that data in order to be able to ask relevant questions. They will need to be familiar with a variety of data tools and analytics models".

Also implied within these research questions is an insightful use of social theory to make better sense of education and digital data. Clearly many of the current concerns and issues associated with the rise of digital data relate to well-established sociological concerns of power, inequality, hierachisation and control from the past one hundred years of social theory. Indeed, there are a number of prevalent theoretical precedents in the emerging literature on digital data and society. Perhaps most obvious are the links back to the work of Foucault – particularly notions of governmentality, biopower, categorization and the Panopticon. Also prominent is Lyotard's writing on performativity, which arose directly from work exploring the rise of computerized systems and databases in higher education. Also of relevance are Durkheim's conceptualization of meaning making, Deleuze's writing on the control society, and Weberian understandings of the rational subject. Similarly, then,

consideration now need to be given to how connections might be made between 'old' theoretical interests from educational studies and the seemingly 'new' concerns of data. These might constructively include Bernstein's notions of code theory, classification and knowledge structures. Also of use is the work of Bourdieu on signs and symbols as mechanisms of power, as well as the transcendental nature of objectivity. As well as drawing on the emerging theorization from the 'digital sociology' literature, there is much that can be brought forward from the past one hundred years of the social study of education.

Making these connections and asking these questions also ties in with a final challenge of formulating alternate conditions for data work – i.e. offering counterpoints to dominant uses of data within educational settings. A starting point here might be to expand upon what Nick Couldry (2013, n.p.) terms 'social analytics' – the study of "how social actors are themselves using analytics - data measures of all kinds, including those they have developed or customised – to meet their own ends, for example, by interpreting the world and their actions in new ways". In this spirit, work can be undertaken that explores how educational research might best support individually-driven and/or community-focussed arrangements for data generation, processing and analysis within educational contexts. Links might be made, for instance, with the growing social movement of 'Open Data'. This concept has developed within information and computer sciences over the past twenty years, advocating unrestricted access and use of 'publically acquired' data for as many people as wish to use it (Gurstein 2011). The practical 'open' uses of digital data has been hastened by the development of increasingly powerful and socially-focused simple software applications that allow non-expert users to directly mine, manipulate and interpret data (Longo 2011). In theory, then, Open Data applications and practices are possible across many areas of education, offering the potential to support radical changes in terms of transparency, accountability, participatory, public engagement and collaborative change.

CONCLUSION

This article has outlined the growing significance of digital data as a topic of interest and enquiry to educational research. While the technical associations of the topic might sit uneasily with their usual concerns of the field, educational researchers and other social scientists with an interest in education should be well placed to provide a necessary critique of digital data. Thus much of what has been argued for in this article stresses the need to recognize – and then act against – the 'politics of data' in education. As Halford *et al.* (2013, p.185) conclude, "it is important that we work to make the social construction of [educational data] visible: to ensure that the micro-politics of its artefacts are understood *as politics*, representing choices and interpretations, rather than as neutral fact or engineering design". Above all, this involves refusing to take digital data 'at face value'. Indeed, this suggests engaging with digital data as much as an 'imaginary' than as a real thing. As Graham and Shelton (2013, p.256) contend, "data is not as an entirely unified and coherent thing around which we should police boundaries but as a set of discourses, objects, and practices". In this sense, the discourses, practices and objects of digital data offer a direct 'way in' to many of the struggles and conflicts that now characterise contemporary education. As such, this is a topic that merits sustained attention across

educational research for many years to come.

ENDNOTE

[1] Anyone writing on this topic faces the issue of whether 'data' is singular or plural (data is *or* data are). I was taught the latter, given the original Latin use of 'data' as the plural of 'datum'. However, with most quoted sources using 'data' as singular, this paper will adopt the same convention to maintain consistency.

ACKNOWLEDGEMENT

REFERENCES

Albrechtslund, A. (2008). Online social networking as participatory surveillance. *First Monday*, 13(3). [firstmonday.org/article/view/2142/1949]

Ayers, D. (2014). When managerialism meets professional autonomy. *Culture and Organization*, 20(2):98-120.

Barnes, T. (2013). Big data, little history. *Dialogues in Human Geography*, 3(3): 297–302.

Beer, D. and Burrows, R. (2013). Popular culture, digital archives and the new social life of data. *Theory, Culture & Society,* 30(4): 47-71.

Bowker, G. and Star, S. (2000). *Sorting things out*. Cambridge MA, MIT Press.

Boyd, D. and Crawford, K. (2012). Critical Questions For Big Data. *Information, Communication & Society*, 15(5): 662-679.

Browne, L. and Rayner, S. (2014). Managing leadership in university reform. *Educational Management Administration & Leadership*, (forthcoming)

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, 31(5): 498-512.

Couldry, N. (2013). *A necessary disenchantment: myth, agency and injustice in a digital world*. inaugural lecture, London School of Economics and Political Science, 21st November

Douglas, J., Douglas, A., McClelland, R. and Davies, J. (2014). Understanding student satisfaction and dissatisfaction. *Studies in Higher Education*, (forthcoming)

Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research?. *Learning, Media and Technology*, 38(3): 237-240

Fenwick, T. and Edwards, R. (2011). Considering materiality in educational policy: messy objects and multiple reals. *Educational Theory*, 61: 709–726.

Fuller, M. (2008). *Software studies: a lexicon*. London, MIT Press

Gandy, O. (2012). Statistical surveillance. In Lyon, D., Haggerty, K. and Ball, K. (eds.). *Routledge Handbook of Surveillance Studies*. New York, Routledge. (pp.125-132).

Geiger, R. and Ribes, D. (2011). Trace ethnography: following coordination through documentary practices. *Proceedings of the 44th Hawaii International Conference*

on System Sciences (pp. 1 -10) [www.stuartgeiger.com/trace-ethnography-hicss-geiger-ribes.pdf]

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, 36(5): 745-766.

Gorur, R. (2011). ANT on the PISA trail: following the statistical pursuit of certainty. *Educational Philosophy and Theory*, *43*(s1): 76-93.

Gorur, R. (2014) Towards a sociology of measurement in education policy, *European Educational Research Journal*, 13(1): 58-72

Gorur, R. and Koyama, J. (2013). The struggle to technicise in education policy. *The Australian Educational Researcher*, 40(5): 633-648.

Graham, M. and Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography* 3(3): 255–261

Grek, S. and Ozga, J. (2010). Governing education through data. *British Education Research Journal,* 36(6): 937-952.

Gurstein, M. (2011). Open data: empowering the empowered or effective data use for everyone?. *First Monday*, 16(2-7) [firstmonday.org/article/view/3316/2764]

Halford, S., Pope, C. and Weal, M. (2013). Digital futures? Sociological challenges and opportunities in the emergent Semantic Web. *Sociology*, 47(1):173-189.

Khalifa, M., Jennings, M., Briscoe, F., Oleszweski, A., and Abdi, N. (2014). Racism? Administrative and community perspectives in data-driven decision making: *Urban Education*, 49(2):147-181.

Kitchin, R. (2014). *The data revolution.* London, Sage.

Kitchin, R. and Dodge, M. (2011). *Code, space: software and everyday life.* Cambridge MA, MIT Press.

Kling, R. (1991). Computerization and social transformations. *Science, Technology and Human Values*. 16(3):342-367.

Knox, D. (2010). Spies in the house of learning: a typology of surveillance in online learning environments. paper presented to *EDGE 2010 - e-Learning: the horizon and beyond* conference, Newfoundland, Canada, October.

Koyama, J. and Menken, K. (2013). Emergent bilinguals: framing students as statistical data? *Bilingual Research Journal*, 36(1): 82-99.

Land, R. and Bayne, S. (2005). Issues of surveillance and disciplinary power in online learning environments. in Land, R. and Bayne, S. (eds). *Education in Cyberspace*. London, Routledge. (pp.165-178).

Laney, D. (2001) '*3D data management: controlling data volume, velocity and variety*' Stamford CT, META Group

Lash, S. (2002). *Critique as Information*. London, Sage.

Latour, B. (2007). Click era spawns a data rich world. in *Times Higher Education*, 6th April [www.timeshighereducation.co.uk/208526.article]

Lingard, B., Creagh, S., and Vass, G. (2012). Education policy as numbers. *Journal of Education Policy*, 27(3): 315-333.

Longo, J. (2011). #OpenData: digital-era governance thoroughbred or new public management Trojan horse?. *Public Policy and Governance Review*, 2(2): 38-51.

Lupton, D. (2013). 'Digital sociology: beyond the digital to the sociological' in *The Australian Sociological Association 2013 Conference Proceedings: Reflections, Intersections and Aspirations, 50 Years of Australian Sociology*. Osbaldiston, N., Strong, C. and Forbes-Mewett, H. (eds). Melbourne: TASA. [www.tasa.org.au/tasa-members/2013-conference-papers/]

Manovich, L. (2011). Trending: the promises and the challenges of big social data. in Gold, M. (ed). *Debates in the Digital Humanities*. Minneapolis MN, University

of Minnesota Press.

Militello, M., Bass, L., Jackson, K. and Wang, Y. (2013). How data are used and misused in schools. *Education Sciences*, 3(2): 98-120.

Monahan, T. (2010). The future of surveillance. *Social Justice,* 37(2–3): 84-98.

O'Keeffe, C. (2014). Assembling the adult learner: transnational and local e-assessment practices. paper presented to *Networked Learning Conference*, April, Edinburgh

Orito, Y. (2011). The counter-control revolution. *Journal of Information, Communication and Ethics in Society*, 9(1): 5-19.

Ozga, J. (2009). Governing education through data in England. *Journal of Education Policy*, 24(2): 149-162.

Ozga, J. (2012) Governing knowledge. *Globalisation, Societies and Education,* 10(4): 439-455.

Palmer, M. (2012). Data flood heightens security fears. *Financial Times*, March 12th [http://www.ft.com/intl/cms/s/0/1aaa2756-58e6-11e1-b9c6-00144feabdc0.html#axzz2otdTNM7u]

Parton, N. (2008). Changes in the form of knowledge in social work. *British Journal of Social Work*, 38(2): 253–269.

Rosenzweig, P. (2012). Whither privacy?. *Surveillance & Society*, 10(3/4): 344-347.

Ruppert, E. (2012). The governmental topologies of database devices. *Theory, Culture & Society,* 29(4/5): 116–136.

Ruppert, E. (2013). Rethinking empirical social sciences. *Dialogues in Human Geography,* 3(3): 268-273.

Savage, M. and Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology,* 41(5): 885-899.

Siemens, G. (2013). Learning analytics: the emergence of a discipline. *American Behavioral Scientist,* 57(10):1380-1400.

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S., Ferguson, R., Duval, E., Verbert, K. and Baker, R. (2011). *Open Learning Analytics: an integrated & modularized platform*. Society for Learning Analytics Research [solaresearch.org/OpenLearningAnalytics.pdf]

Tække, J. (2011). Digital Panopticism and organizational power. *Surveillance and Society*, 8(4): 441-454.

Taylor, E. (2013). *Surveillance schools*. Basingstoke, Palgrave Macmillan.

The Economist (2010). The data deluge. *The Economist*, February 25th [www.economist.com/node/15579717]

Thrift, N. (2005). *Knowing capitalism*. London, Sage.

Webster, A. (2013). Digital technology and social windows. in Orton-Johnson, K. and Prior, N. (eds). '*Digital sociology*'. Basingstoke, Palgrave (pp.227-233)

Yoo, Y., Boland, R., Lyytinen, K. and Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science* 23(5): 1398-1408.