ESSAY COVER SHEET 2018

NIR605: Critical Data Studies

**Instructions:**

All work must be typed and pages must be numbered. This cover sheet must be clearly visible at the front of all work.

Full Name: **Sean O'Riogain**

Student Number: **18145426**

Date Submitted: **8th March 2019**

Course Title & Lecturer: **Critical Data Studies (NIR605), Dr. Rob Kitchin.**

Essay/Project title: **A Critique of the Paper 'What Makes Big Data, Big Data?' by Kitchin R. & McArdle G. (2016).**

# NIR605 – Critical Data Studies - Assignment 1

## A Critique of the Paper 'What Makes Big Data, Big Data?' by Kitchin R. & McArdle G. (2016)

### Overview

#### Context

In this paper, Kitchin et al. contend that the level of ontological analysis of the phenomenon that is commonly known as 'big data' has not been done to a sufficiently low level of detail to enable the term to be properly understood. It acknowledges that, since it was coined by John Mashey in the 1990s, its ontology has evolved to include characteristics such as volume, velocity and variety - as encapsulated by the '3 Vs' meme that emerged from the early work of Doug Laney (2001) - as well as resolution, relationality and indexicality, and flexibility and scalability more recently, which the paper refers to collectively as Kitchin's taxonomy, and that other, albeit non-ontological, equivalents have arisen from the work of researchers such as Uprichard (2013) and Lupton (2015) in the social sciences.

#### Objective

However, the authors posit that these features need to be looked at more closely to be able to validate them and to rank them in terms of the importance of the contribution they make to the essence of the big data concept. They argue that this work is an important prerequisite for positioning future epistemological studies in the big data field.

#### Methodology

The authors undertook this study by reviewing 26 different types of dataset, selected from 6 different domains, that are commonly considered to be in the big data category and with which they have had some experience, using the 7-element taxonomy referred to above.

#### Findings

The main conclusion of this study of 26 types of big data set types was that very few of them exhibited all 7 characteristics and some of them did not even include volume (which implicitly challenges the use of the adjective 'big' in 'big data'). In contrast, the authors found that velocity and exhaustivity were the properties that were most common across the datasets. This insight led them to conclude that those 2 features were the ones that were essential to understanding the nature of 'big data' as a concept.

Other valuable, if secondary, points raised by this study was that volume was typically a by-product of velocity and exhaustivity (and, therefore, is derivative of them) and that there are two types of velocity – that at the point of capture of the data and during its subsequent processing – which need to be considered separately. By the same token, two other elements of the relevant taxonomy (resolution & indexicality and extendibility & scalability) also need to be looked at under their component parts as they are not synonymic. Finally, the authors also point out that the volume characteristic of big data sets is typically measured in terms of the total storage capacity needed to accommodate it (e.g. in terms of terabytes, petabytes etc.) and that, by not considering it in its two dimensions –in terms of both breadth (columns) and depth (rows) – separately, we run the risk of misclassifying a data set as something other than big data when it actually belongs in that category.

# NIR605 – Critical Data Studies - Assignment 1

## Critique

In her call to arms to her fellow social scientists in relation to big data, in her blog, Emma Uprichard (2013) bemoaned the use of the adjective 'big' when she remarked that "With all of the data power it [big] is supposedly meant to entail, one might have thought that a slightly better descriptive term might have been latched onto. But, no. BIG. Just BIG." Had those who coined the term 'big data', and who subsequently propagated its use, been in possession of this paper, her wish may well have been granted – especially in the light of how it credibly argues for the downgrading of the importance of its volume attribute.

While care has obviously been taken to ensure that the number and variety of the big data sets that are considered using this paper's methodology are sufficiently extensive, a clear pattern is visible in the two first columns of the table (3) in which they are analysed; they refer to social interactions almost exclusively – from communicating via mobile telephony, social media and other website, to building, buying and improving real estate, seeking employment, and traveling by road and air etc. Of course, this supports Uprichard's assertion (in the same blog) that "After all, most big data is and will continue to be social data."

Having said that, it is doubtful as to whether the data set types referred to in the Administrative category qualify as big data because, while they clearly strive for exhaustivity, they do not possess velocity at any stage during their respective lifecycles - and especially in light of the fact that the paper eliminates another, and arguably similar, type of dataset (i.e. census data) from the big data category. On the other hand, including those data sets does provide some spread across the private and public sectors. They also play a part in diluting the importance of the volume attribute.

For future research, it would be valuable to extend this methodology from the general (generic) to the specific in terms of the big data sets under analysis. As this would require inter-disciplinary collaboration (especially with computer science) and the active support of suitable enterprises in both the public and private sectors, this could turn out to be a major undertaking. However, it could provide a powerful vehicle for validating and enhancing both the methodology and the results that it achieved - including, for instance, improving the definition of the metrics used to assess the big data sets under each of the headings.

While this paper does not give rise to a meme of its own – like the '3 Vs' or the '13 Ps' – it does make use of another device for knowledge encapsulation, dissemination and retention – the analogy or metaphor – which is powerfully used to underline one of its main insights: that the big data 'genus' is comprised of multiple 'species' based on the their profile vis a vis Kitchin's taxonomy. As Puschmann and Burgess (2014) state in their in their treatise on the use of metaphor in relation to big data: "Metaphor, in the cognitive view, is an important conceptual tool that enables us to understand abstract concepts in terms of more familiar and concrete ones" (p. 1695). Ironically, memes themselves are thought to exhibit Darwinian properties that dictate their propagation and, ultimately, their survival. Notwithstanding the fact that this paper downgrades the importance of the volume characteristic of big data, it is unlikely that the 3-Vs meme will become extinct in the foreseeable future as it still refers to 3 of the 7 (or 9) elements of Kitchin's taxonomy.

In conclusion, and notwithstanding the previous (relatively incidental) observations, this paper provides some key new insights in relation to the ontology of big data, and provides a launch pad from which previous studies in this area – both ontological and epistemological - can be revisited/reassessed and from which new explorations of the topic can be undertaken more confidently, armed with what will undoubtedly become a common view of what big data essentially are.

## References

Kitchin, R. and McArdle, G. (2016) What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society* 3: 1–10. Available at: https://journals.sagepub.com/doi/abs/10.1177/2053951716631130 (accessed 8 March 2019).

Laney D. (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed 8 March 2019).

Lupton D. (2015) The thirteen Ps of big data. *The Sociological Life*, 13th May. Available at: https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-data/ (accessed 8 March 2019).

Puschmann C. and Burgess J. (2014) Metaphors of Big Data. *International Journal of Communication* 8: 1690–1709. Available at: https://ijoc.org/index.php/ijoc/article/view/2169/1162/ (accessed 8 March 2019).

Uprichard E. (2013) Big data, little questions. *Discover Society*, 1st October. Available at: https://discoversociety.org/2013/10/01/focus-big-data-little-questions/ (accessed 8 March 2019).