

# How has the status of data in society changed in recent years?

## Introduction

In recent years, data has become a mainstream concept, whether the premise of sci-fi movies such as *The Matrix* and *Minority Report*, animated visualisations on the BBC's coverage of election night or stylised infographics in *The Guardian* newspaper. Data are both generated and utilised by digital services, including retail websites, social media services and location-aware mobile devices. In an increasingly datafied world, the recognised basis of data as a technological component of digital services deliberately hides the true status of data as a new yet largely invisible form of power. As Turow describes it, "We're at the start of a revolution in the ways marketers and media intrude in and shape our lives" (2011, pp2).

This essay begins by providing a definition of data in the context of answering the question. To understand how the status of data in society has changed in recent years, a short history and background to data is presented to consider how data has changed. The subsequent section explores how the status of data was significantly impacted by the advent of the social web and the arrival of 'big data', and how data empowers knowledge generation through the use of algorithms. Finally, in conclusion, brief consideration is given to some of the challenges for data moving forward and whether these might affect the status of data in future.

## A background to data

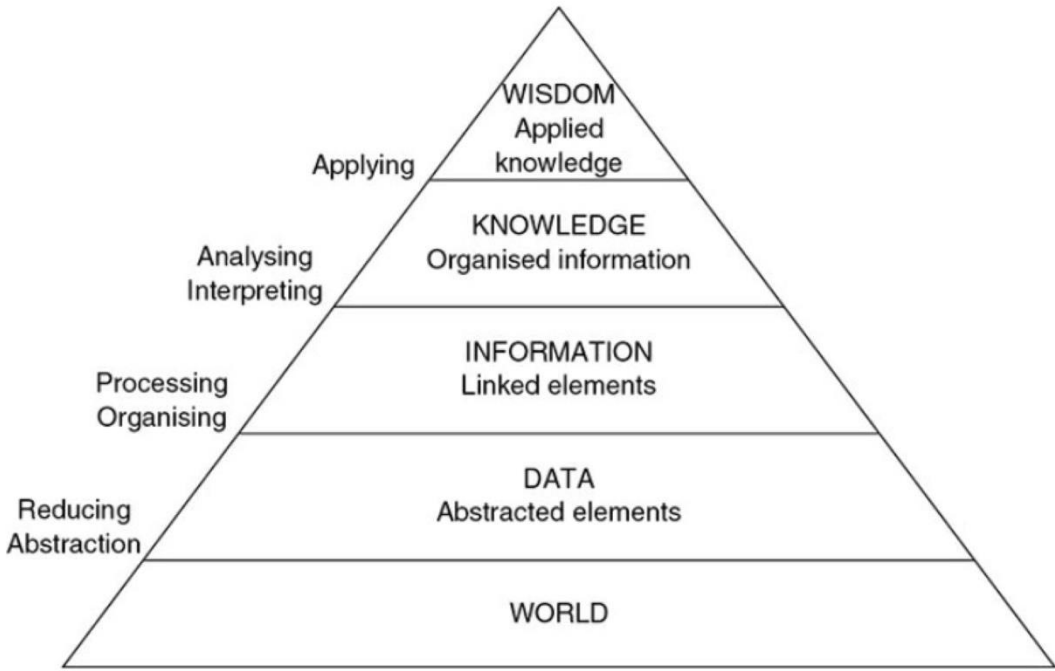
In this section I will provide a definition of data as it used in this essay and, to subsequently consider how the status of data has changed, I will look briefly at the history of data.

Data can be produced as an abstracted outcome of describing, measuring or observing the world, resulting in a comparable, correlatable and quantifiable representation, such as in the form of numbers, categories or text (Kitchin 2014, pp2). Datafication is the process of taking anything we can describe, measure or observe and turning it into quantifiable data (Mayer-Schönberger and Cukier, 2013, pp15). Each individual quantifiable attribute captured is a datum, though generally the term data is used whether referring to a single

data-point or many. Furthermore, data frequently comes with attached metadata such as additional information which extends the description of each individual data-point or of the entire dataset (Kitchin 2014, pp8-9).

Data can be highly useful, and therefore is often considered valuable, because it enables the automated processing and analysis of the world. More explicitly, a widely recognised model (Figure 1) describes how data can be used to generate information, and in turn, can reveal knowledge and generate new wisdom when applied. Data can therefore be regarded as a fundamental component of how we make sense of the world, and as such is considered highly valuable and of high status (Kitchin 2014, pp9-12). Hence, when considering the status of data, this essay is concerned with the power that data provides through abstracting elements, revealing and creating knowledge, and hence creating an opportunity to monetise data.

Figure 1. Knowledge pyramid (Kitchin 2014, pp9)



Until the early 21st century, use of the word data primarily related to the storage of computer data, such as hard-drives on personal computers or larger-scale data storage for storing information on magnetic tape or hard-drives (Kitchin 2014, pp85). However, with an increasingly wider availability of the internet from the mid 1990s onwards, the need for advertisers to monitor the performance of adverts on websites lead to an early form of online data collection. With far more accuracy than traditional print or television advertising, advertisers could track individual user’s clicks, which showed which adverts

they did (or didn't) interact with. The tracking of individual clicks subsequently evolved to 'cookie' technology, which enabled websites to store multiple items of small structured information in the user's browser. This further allowed for continuity between visits to track many aspects of online behaviour such as products viewed, abandoned carts or completed purchases (Turow 2011, pp47-48). Many current tracking and data collection methods still rely on cookie technology.

The following section considers the major shift in the status of data with the advent of the social web, the widespread industry adoption of Big Data and the use of algorithms to generate knowledge from data.

## The dramatic change in data's status

In this section, the seismic shift in the status of data is considered with regards to two related factors – the advent of the social web, often referred to as 'web 2.0', and an industry-wide focus on Big Data – and the utilisation of algorithms to enable data to generate knowledge.

### Web 2.0 – the social web

The nature of what users do on the internet changed dramatically with the advent of 'web 2.0', a term which gained huge traction around 2004 to describe a new web experience (O'Reilly 2005, Section 1). Technological definitions of web 2.0 focus on the advance in web browsers and programming technologies which enabled richer, interactive website experiences combined with the greater bandwidth available with broadband internet. More significantly though, sociological and conceptual definitions of web 2.0 recognise a massive shift in how it changed what users do on the internet. Until this point, the internet was predominantly a broadcast medium requiring technical expertise to build and maintain websites which enabled users to consume information or purchase products. Web 2.0 created a more collaborative, self-publishing mode of interaction and engagement with the internet, including blogging, commenting, uploading photos and the sharing of 'real-life'; what would become recognised as social media (Kitchin 2014, pp99).

Where using websites had previously been about 'browsing', the new language of web 2.0 focussed on user participation through the production or sharing of content, making connections or engaging with communities. This re-positioned the web as a rapidly expanding social medium (Beer 2009, pp986; van Dijck 2013, pp6-14). Indeed, even the term 'website' evolved such that web experiences were less about single destination

websites and more about personalised interactive web services, built upon the interconnected web 2.0 infrastructure (van Dijck 2013, pp6). From the perspective of the social network site owners, this new paradigm for the web generated a new, largely unforeseen but predictably valuable commodity through the datafication of interests, web searches, comments, user-generated content and connections (van Dijck 2014, pp198). Whilst users were exploring the opportunities and experiences of a more interactive and potentially engaging internet, the social media companies identified “an investor-oriented line between the burst Internet bubble of 1999 and a future that might again make money” (Baym 2015, pp1).

## Big Data

At about the same time as web 2.0 changed the experience of using the internet, computing power and data capture technology evolved into what is now regarded as ‘Big Data’. Whilst the term Big Data had limited use as early as the 1990s in relation to scientific analysis of huge datasets (Kitchin 2014, pp67), it wasn’t until approximately 2010 that the term “acquired such popular resonance” (Crawford et al 2014, pp1664), prompting boyd and Crawford to declare that “the era of Big Data has begun” (2012, pp662). Whilst there is no explicitly agreed definition of Big Data, Kitchin provides a detailed summary of the widely accepted understanding based on the three Vs – volume, velocity and variety. Volume expects the data to be huge, requiring measurement in at least terabytes (1,000 gigabytes) or petabytes (1,000 terabytes). Velocity expects the data to be current, typically created in near real-time. Variety expects the data to be diverse, whether structured or unstructured, qualitative or quantitative (2014, pp68). More recently, this common understanding of the three Vs has been expanded to reflect the more advanced practices that are now common. Some researchers have added more Vs such as *Value* to represent the inherent financial value and *Variability* to reflect the expected statistical variance (Esposti 2014, pp2). Alternative additions include being *Exhaustive*, based on the increasing practice of capturing all data, all of the time; and *Flexible*, based on recent database approaches to store data in less rigid, less structurally constrained ‘NoSQL’ databases (Kitchin 2014, pp69). Other definitions of Big Data worry less about describing exactly what it is and more about what can be done with it, advocating alternative understandings such as the ability to interrogate it, analyse it and cross-reference it with other data (boyd and Crawford 2012, pp663). Furthermore, other less-explicit definitions consider what Big Data is more generally about, such as Mayer-Schönberger and Cukier’s

who posit that “big data is all about seeing and understanding the relations within and among pieces of information that, until very recently, we struggled to grasp” (2013, pp19).

In the last few years, as collecting and analysing Big Data has become a standard mode of operation, usage of the term has already begun to diminish (Kennedy 2016, pp10).

However, the power of Big Data analysis only continues to increase as more data becomes available to correlate with other data. In 2013, the CIA's Chief Technology Officer Gus Hunt suggested that the true power of data lies in its correlation with other data.

Therefore, as the value of any information is only known when it is connected with possibly as yet uncollected information, the only sensible approach is to collect and store everything, all of the time (Sledge 2013). This suggests that two potential limitations to Big Data - scale of storage and computational power – have been overcome, and therefore the status of data will continue to increase. Further, commentators such as Anderson (2008) have proposed that this naturally leads to the end of theoretic, qualitative analysis as “with enough data, the numbers speak for themselves” and hence the need for academic theories and qualitative analyses are rendered redundant. Mayer-Schönberger and Cukier claim that correlations are preferable to causation, and that “for many everyday needs, knowing what not why is good enough” (2013, pp191).

## Algorithms

Whilst web 2.0 and Big Data have established the foundation of data's increasingly high status, the role of the algorithm further enables data to enact its powerful status because, as Goffey states, “algorithms do things” (2008, pp17). Algorithms may simply be thought of as “recipe-like sets of instructions to solve problems” (Pasquale 2015, Para 4). They can range from simple to vastly complex and consist of a set of operations to be performed, potentially including rules which may trigger further operations (Goffey 2008). Algorithms significantly change how meaning and knowledge are generated because rather than applying expertise and selective methodologies to analyse data, algorithms programmatically generate knowledge from the data (Kitchin 2014, pp2). In conjunction with advances in computing power and the capacity to handle vast, unstructured datasets, algorithms can react in real-time to find new ways of generating knowledge by identifying patterns in the data (Mcquillan 2015, pp567). Furthermore, as well as data thereby enabling the algorithmic generation of knowledge about past events, it can be “analyzed and processed into predictive algorithms about future human behavior” (van Dijck 2014, pp201). Lash asserts therefore that we now live in “a society in which power is increasingly in the algorithm” (2007, pp71), which Beer summarises as having “the capacity to shape

social and cultural formations and impact directly on individual lives” (2009, pp994). Data can therefore be seen to empower what Striphas calls the “algorithmic culture”, whereby data can be considered to encapsulate our thoughts, conduct and expressions (2015, pp396).

## The high status of data

Recognising the introduction and continued growth of the social web since 2005 in conjunction with the huge adoption of Big Data approaches and algorithmic analysis since 2010, it becomes clear that the status of data has changed significantly in recent years. Through datafication, social media services can now transform everyday social interactions into data (Mayer-Schönberger and Cukier 2013, pp91), creating a “new paradigm for understanding sociality and social behavior” (van Dijck 2014, pp198). Considering this transformation of social engagement and reflecting on Figure 1 from the previous section, it is apparent that data has acquired high status due to its broad utility in generating knowledge. Knowledge has high value for platform owners in enhancing their understanding of users’ behaviour and evolving their own service, but potentially of more significance is the value of this type of knowledge to advertisers. Turow (2011) explains:

“Advertisers in the digital space expect all media firms to deliver to them particular types of individuals - and, increasingly, particular individuals - by leveraging a detailed knowledge about them and their behaviors that was unheard of even a few years ago.” (pp4)

In real terms, the more accurately an advert can be targeted, which is increasingly at “unprecedented resolution” (John 2016, pp65), the higher the price advertisers will pay to place their advert.

Within social media, arguably the most successful company at leveraging the power of data to generate highly valuable knowledge for advertisers has been Facebook, generating \$10 billion from ad revenue in Q3 2017 (BBC 2017, Para 1). Yet Facebook’s vast financial success, one of only five companies in the world to be valued at more than \$500 billion (Egan 2017, Para 2), is built on a model which is deliberately kept invisible from its users as it recognises that “unveiling too many new purposes for its users’ data too soon could freak them out” (Turow 2011, pp96). Until recently Facebook described its mission as being “to make the world more open and connected” (John 2016, pp65). Yet it is far from transparent and open about its own role in datafying user content and behaviour

to sell user profiles to advertisers, prompting Baym (2015) to describe the process as “the takeover of the social by the corporate” (pp1). Facebook and other social media services have deliberately evolved a concept of sharing on social media as a state of connectedness to others (John 2016, pp55), yet “it serves to paper over the commercial aspects of the ways in which many Social Network Services operate” (John 2016, pp64). Furthermore, by developing the premise of sharing from an idea of “forming and maintaining intimate relations” (John 2016, pp10), it could be considered more intrusive that the data users are encouraged to generate on social media is intrinsically more personal to them than perhaps banking or shopping data. Despite this, the unseen datafication of social media content and behaviour generates the most lucrative of data (Kennedy 2016, pp25).

In the following concluding section, considering the high status data has attained and is increasing with few limitations, thought is given to the possible future challenges for our society and whether these might affect the status of data in future.

## Considering the future of data and conclusion

In this final section, brief consideration is given to potential future challenges from data having attained such high status before concluding by summarising how the status of data has changed in recent years.

### Future considerations regarding the status of data

The power of data is likely to increase as more data is collected, not only as the generation new data has inherent value but as previously unutilised data may become useful through new correlations. On this trajectory, data will continue to generate more wealth, as John asserts, “... every time we share something online, we create traces of data... the more we share online, the richer the online platforms we are sharing on become” (2016, pp66). The economic motives of profit-driven private business are likely to have differing objectives of data mining to the individual user, many of whom may not even be aware of these practices, let alone have an ability to impact on them (Bollier 2010, pp23). However, even in specific cases it may be too simplistic to consider data as being used in good or bad ways; the aims and outcomes of data usage are a complex mix of personal benefits and corporate or state objectives. For example, the vast use of data for varied intentions may include aims for safer, more productive or accountable societies yet do so by monitoring, observing and analysing individuals (Kitchin 2014, pp165). Pasquale posits

that regulation is required in order to “help consumers recognize the perils of the new information landscape without being overwhelmed with data.” (2014, Last para).

From a broader perspective, it can be considered that datafying humans as profiles for advertisers to buy is at odds with using technology to create a better society (Baym 2015, pp2). Whilst posing their “Critical questions for Big Data”, boyd and Crawford (2012) illustrate the classic dystopian picture of a Big Brother surveillance state before describing a series of specific challenges for Big Data in the future, covering issues such as consent, privacy, anonymity and ownership. Kitchin (2014) concludes his assessment of the wider implications of data changing society by declaring it “vital that scholars take an active role in researching and thinking through the ethical, social, political and legal questions arising from the data revolution” (pp183). Similarly, in concluding their relatively positive stance on the Big Data revolution, even Mayer-Schönberger and Cukier warn that “we must use this tool with a generous degree of humility... and humanity” (2013, pp197).

## Conclusion

In considering how the status of data has changed in recent years, it is evident that data today is vastly more powerful than even a decade ago. In defining what data is, it is apparent through models such as Kitchin’s “knowledge pyramid” (2014, pp9) that data is fundamentally powerful due to its role in generating knowledge. The evolution of the internet with the advent of web 2.0 (O’Reilly 2005) created a shift to web services and platforms which established the web as a social medium (Beer 2009; van Dijck 2013), creating an explosion in data generation and recreating an essential opportunity for social media platforms to generate revenue (boyd and Crawford 2012, pp662; Baym 2015, pp1). As computing power and storage capacity evolved, significant growth and adoption of utilising Big Data approaches further increased the status of data, as observed by Mayer-Schönberger and Cukier in recognising that Big Data enables us to generate knowledge that previously we couldn’t identify (2013, pp19). Furthermore, the status of data is further increased when algorithms are applied, enabling data to empower the generation of knowledge in real-time despite increasingly adapting to new and vast collections of unstructured, varied data (Mcquillan 2015, pp567). Data’s high status is particularly appreciable when recognising not only the ability to generate knowledge of historical events but increasingly in generating knowledge through predicting future outcomes (van Dijck 2014, pp201). The vast power and financial success of companies such as Facebook can be attributed largely to the application of datafication processes, generating vast advertising revenues by using data to understand social interactions and social behaviour



(Mayer-Schönberger and Cukier 2013, pp91; van Dijck 2014, pp198). The lack of visibility and transparency in much of this datafication enables the status of data to remain largely unthreatened despite being seen as “the takeover of the social by the corporate” (Baym 2015, pp1). Issues affecting the status of data such as consent, privacy, anonymity and ownership will be the subject of further research by academics and may increasingly be the subject of users’ attention (boyd and Crawford 2012). The status of data remains unlikely to be affected whilst private businesses use datafication to deliver against profit-driven objectives, leaving users unaware of these practices and without an ability to impact on them (Bollier 2010, pp23). It is apparent that whilst all aspects of society including how we live, work and think, are impacted by the power of data (Mayer-Schönberger and Cukier, 2013, pp190), data will continue to attain high status. Regarding how the status of data has changed in society in recent years, Kennedy succinctly asserts, “it is not data’s size, but its power that matters in contemporary society” (2016, pp10).

## References

- Anderson, C. (2008). The end of theory: Will the data deluge make the scientific method obsolete? Edge. Available at: [http://www.edge.org/3rd\\_culture/anderson08/anderson08\\_index.html](http://www.edge.org/3rd_culture/anderson08/anderson08_index.html) (Accessed 16/12/17).
- Baym N. (2015) ‘Social Media and the Struggle for Society’, *Social Media + Society*, Volume 1 (Issue 1). Available from: <http://journals.sagepub.com/doi/full/10.1177/2056305115580477> (Accessed 12/12/17).
- BBC (2017) Facebook ad revenue tops \$10bn. Available at: <http://www.bbc.co.uk/news/business-41835989> (Accessed 17/12/17).
- Beer D. (2009) ‘Power through the algorithm? Participatory web cultures and the technological unconscious’, *New Media & Society*, Volume 11(6), pp985-1002.
- Bollier, D. (2010). The Promise and Peril of Big Data. Available at: <http://www.aspeninstitute.org/publications/promise-peril-big-data> (Accessed 15/12/17).
- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon, *Information, Communication and Society*, Volume 15 (Issue 5), 662–679. Available at: <http://www.tandfonline.com/doi/full/10.1080/1369118X.2012.678878> (Accessed 05/12/17).
- Crawford K., Miltner K. and Gray M. (2014) ‘Critiquing Big Data: Politics, Ethics, Epistemology’, *International Journal of Communication*, Volume 8, pp1663–1672. Available from: <http://ijoc.org/index.php/ijoc/article/view/2167/1164> (Accessed 13/12/17).

- Egan M. (2017) Facebook and Amazon hit \$500 billion milestone. Available at: <http://money.cnn.com/2017/07/27/investing/facebook-amazon-500-billion-bezos-zuckerberg/index.html> (Accessed 15/12/17).
- Esposti S. (2014) 'When big data meets dataveillance: The hidden side of analytics', *Surveillance & Society*, Volume 12 (Issue 2), pp209-225. Available from: <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/analytics> (Accessed 16/12/17).
- Goffey, A. (2008). Algorithm. In M. Fuller (Ed.), *Software studies - A lexicon* (pp. 15–20). Cambridge: MIT Press.
- John N. (2016) *The Age of Sharing*, Cambridge, Polity Press.
- Kennedy H. (2016) *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*, London, Macmillan Publishers Ltd.
- Kitchin R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*, London, SAGE Publications Ltd.
- Lash, S. (2007). Power after hegemony: Cultural studies in mutation. *Theory Culture & Society*, Volume 24 (Issue 3), pp55–78. Available at: <http://journals.sagepub.com/doi/pdf/10.1177/0263276407075956> (Accessed 12/12/17).
- Mayer-Schönberger V. and Cukier K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*, London, John Murray Publishers.
- McQuillan, D. (2015) 'Algorithmic states of exception', *European Journal of Cultural Studies*, Volume 18 (4–5), pp564–576. Available at: <http://journals.sagepub.com/doi/full/10.1177/1367549415577389> (Accessed 16/12/17).
- O'Reilly T. (2005) *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Available at: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> (Accessed 17/12/17).
- Pasquale F. (2014) *The Dark Market for Personal Data*. Available at: <https://www.nytimes.com/2014/10/17/opinion/the-dark-market-for-personal-data.html> (Accessed 18/12/17).
- Pasquale F. (2015) Algorithms are producing profiles of you. What do they say? You probably don't have the right to know. Available at: <https://aeon.co/essays/judge-jury-and-executioner-the-unaccountable-algorithm> (Accessed 18/12/17).
- Sledge M. (2013) CIA's Gus Hunt On Big Data: We 'Try To Collect Everything And Hang On To It Forever'. Available at: [http://www.huffingtonpost.co.uk/entry/cia-gus-hunt-big-data\\_n\\_2917842](http://www.huffingtonpost.co.uk/entry/cia-gus-hunt-big-data_n_2917842) (Accessed 16/12/17).
- Turow J. (2011) *The Daily You: How the new advertising industry is defining your identity and your worth*, London, Yale University Press.

van Dijck, J. (2013) *The culture of connectivity: A critical history of social media*, Oxford, Oxford University Press.

van Dijck, J. (2014) 'Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology', *Surveillance and Society*, Volume 12 (Issue 2), pp197–208.

Available at: <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/datafication> (Accessed 12/12/17).