Article

Tales of the digital sublime: Tracing the relationship between big data and professional sport

Brett Hutchins
Monash University, Australia

Abstract

This article outlines the relationship between Big Data and sport in the network society. Critiquing the hype associated with Big Data, it is explained that modern sport informs the historical rise of this technological phenomenon, serving as a social and cultural site where the accelerating privatization and commodification of statistics and statistically generated information occurs. These developments deliver increased entertainment options for fans of many professional men's sports and an unprecedented number of performance indicators for selected coaches, athletes and pundits. However, the information technology infrastructure and resources required to generate real-time data are adding to widening inequalities between elite 'data-rich' sports and comparatively impoverished 'data-poor' sports, including many women's competitions. It is argued that a collective fascination with the digital sublime obscures the complex interaction between corporate power, digital data markets, history and culture, and contributes to inequalities that demand ongoing attention and critique.

> The fashionable term now is 'Big Data'. IBM estimates that we are generating 2.5 quintillion bytes of data each day, more than 90 percent of which was created in the last two years. (Silver, 2012: 9)

> The market sees Big Data as pure opportunity: marketers use it to target advertising, insurance providers use it to optimize their offerings, and Wall Street bankers use it to read the market. (boyd and Crawford, 2012: 664)

Corresponding author:
Brett Hutchins, School of Media, Film and Journalism, Monash University, PO Box 157, Caulfield East, Melbourne, VIC 3145, Australia.
Email: brett.hutchins@monash.edu

> [Data] is the most valuable asset, being the image of the world in which we are attempting our endeavours. Prizing data is something Google understood from the start. (Dumbill, 2013: 71)

Data are viewed as a source of significant market power in the network society, with the production, control and sale of large-scale digital data being the subjects of increasing enthusiasm across an array of social, political and economic contexts. As the opening quotations indicate, the prizing of data and the emergence of predictive analytics are creating new markets and commercial opportunities, with information technology and digital media behemoths such as IBM and Google cashing in on the exabytes of data generated each day. Big Data[1] has caught the attention of powerbrokers in almost every conceivable area of social, cultural and economic activity, including the sciences, the arts, government, education, health care, human resource management, gambling, the insurance industry, the military, intelligence services and policing, traffic planning, retailing, product design and marketing. These developments are both exciting and unnerving, focussing attention on the turbocharged processing capabilities and voluminous storage capacity of computing technologies, but also downplaying the inequalities and unintended consequences of pervasive digital technology infrastructures and data collection.

Scepticism, context and historical perspective are essential tools to employ when faced by the seemingly unstoppable ascendancy of Big Data. It is time to slow down and ask critical questions about the practices and structures that intersect with Big Data (boyd and Crawford, 2012; Gane, 2006), especially given that it signals the proliferation of historically unprecedented computational power, networking scale and data volume in a variety of settings. These developments demand analysis of the dynamics of data production and access, as well as the potential outcomes and risks flowing from the questionable assumptions and biases related to Big Data (boyd and Crawford, 2012: 665; Crawford, 2013). This article addresses these issues by analysing a specific area of popular media and social activity, sport. The rise of Big Data, and the market and media discourses surrounding it, connects to an expansion in the quantification of performance measures throughout the history of modern sport. In turn, this history speaks to a long-standing pattern where human endeavours have been subject to 'data-isation' (Hunter, 2006) that predates the Internet, mobile media and the era of Big Data. The advent of digital networked communications and technologies, and the capacity for real-time data generation, is simultaneously solidifying these practices and altering the patterns associated with them.

The data harvested by the measurement and quantification of sport, as well as the activities of fans and consumers, continue to grow. As will be shown, these practices can be understood via mechanisms for the *generation* of data, the *value* assigned to data sets produced by the control of costly computing technologies and digital media infrastructure, and the commercial opportunities and services created by the capacity to *access* data. The picture that emerges builds on the arguments of Mark Andrejevic (2013) about the state of culture and politics in an age of 'infoglut':

> At stake ... is a reconfiguration of the relationship between forms of knowledge and power. Two aspects of this relationship are of particular concern: first, the increasing asymmetry between those who are able to capture, store, access, and process the tremendous amounts of data produced by the proliferation of digital, interactive sensors of all kinds; and, second, ways of understanding and using information that are uniquely available to those with access to the database. (p. 17)

The rewards for those existing on the privileged side of the 'data analysis divide' (Manovich, 2012: 461) are market power and profits for already wealthy and heavily covered men's sports leagues and events. For those sitting on the wrong side of this divide, including many women's

sports, the seemingly intractable inequalities of television coverage, news reporting and sponsorship investment that took root in the age of mass media are continuing (Communication & Sport, 2013; Hargreaves, 1994), particularly as the collection and uses of data expands each year. Yet, these existing conditions and unfolding inequalities are obscured by a popular fascination with new digital media technologies, demonstrating the continuing influence of 'the digital sublime' in late capitalist societies (Mosco, 2004).

Big Data and the digital sublime

The excitement surrounding new data applications is producing a cacophony of claims that make it difficult to establish a clear and agreed upon definition of Big Data. These claims are fed into a marketplace for citable neologisms such as 'datafication' (Lycett, 2013) and 'infobesity' (Mouncey, 2012), conveying the idea of a new frontier of innovation. The following statement about the Big Data phenomenon undercuts generalizations and futurist abstractions by emphasizing the *relative* connotations of the term: 'The term *Big Data* has a relative meaning and tends to denote bigger and bigger data sets over time' (Mahrt and Scharkow, 2013: 22; original emphasis). A relatively large scale and volume of digital data now resides in many privately owned and, to a lesser extent, publicly accessible servers connected by distributed computing, including cloud and grid systems. Data are processed by algorithmic software that collates, compares, recombines and analyses sections of data to achieve designated and unforeseen insights. Thus, Big Data is not just about size, but 'a capacity to search, aggregate, and cross-reference large data sets' (boyd and Crawford, 2012: 663). It is in this *capacity*, and the expanding number of analytical potentialities on offer, that the alchemy of Big Data can be found. A diverse collection of organizations are discovering new ways to gather data (such as smartphones, tablets, apps, wearable biofeedback devices, social networking services and locative media), as well as identifying unfamiliar facts in existing and combinable data sets after subjecting them to algorithmic analysis (including census data, purchasing records, health and government records and voting records). The design and control of the formulas and methods used for the collection, organization and analysis of Big Data are key to the extraction of commodity value in a burgeoning data marketplace.

Purveying its typical mix of insight and digital culture boosterism, *Wired* has been publishing stories about Big Data complete with captivating infographics. Big Data is, we are told, an 'information revolution' that exists 'at an almost unimaginable scale' (Pearlstein, 2013), 'a golden age for as-it-happens actionable insights' (Lorentz, 2013) and a technology for a 'new world order' (McMillan, 2013). These declarations and their accompanying stories reveal the excitement that is attached to the immense scale, volume and diversity of stored and retrievable digital information, indicating that Big Data is about much more than technology. As danah boyd and Kate Crawford (2012) identify, Big Data rests on the interplay between technology, analysis and a *mythology* that triggers both utopian and dystopian rhetoric (p. 663). The final component of boyd and Crawford's definition pairs 'cutting-edge' technology with deeply rooted historical and cultural patterns that, in this case, contribute to the exultation of technologies that promise to deliver epistemological breakthroughs, '*Mythology:* the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (p. 663; original emphasis).

This 'higher' form of intelligence devalues individual comprehension in favour of database technologies that are able to compensate for the supposed 'shortcomings of the human brain by ushering in new forms of aggregate "knowledge" and predictive power' (Andrejevic,

2013: 20–21). Big Data is made synonymous with fact in this formulation, constituting a 'gift' that obliterates the problem of human subjectivity and resolves once unanswerable questions (Gregg, 2015).

Big Data is embedded in a much older story about the cultural reception of new media and communications technologies, including the telegraph, telephone, television and the Internet, which often masks their military and industrial origins (Mosco, 2004). The mythologizing process under consideration here functions in two ways (cf. Barthes, 1973; Carey, 1989; Carey and Quirk, 1970; Denzin, 1987; Maxwell and Miller, 2012; Miller, 2012; Mosco, 2004). First, the popular discourses used to promote 'exciting' new technologies emphasize their value through an appeal to deep-seated emotions such as desire, hope and wonder. Often elicited through sophisticated advertising and public relations campaigns, these feelings promise future prosperity, wellbeing and a 'higher form of intelligence and knowledge' (boyd and Crawford, 2012: 663). Second, the social, political, legal and ecological impacts and risks flowing from the introduction of these technologies are depoliticized and subsumed under the signs of progress and future-oriented innovation. These signs de-emphasize the vested interests wanting to popularize and profit from the technology and/or recast these interests as visionaries and innovators. This recasting also assists in clearing the way for politically palatable government subsidies, both direct and indirect, that are thought essential for the health of advanced economies.[2]

Early signs of the mythologizing of Big Data can be found in the speculations of Marshall McLuhan during the 1960s. McLuhan noted the conjunction between the growth of mainframe computing and the collection of data by corporations, market research firms and the social scientific community:

> We have reached a similar point when each stick of chewing gum we reach for is acutely noted by some computer that translates our least gesture into a new probability curve or some parameter of social science. Our private and corporate lives have become information processes just because we have put our central nervous systems outside us in electric technology. (McLuhan, 1964: 51–52)

McLuhan is followed in 1970 by James W. Carey and John J. Quirk (1970; Carey, 1989) who combine cultural analysis and historical sensibility to identify the relationship between new electrical technologies and a 'rhetoric of the technological sublime' during the 19th and early 20th centuries (p. 395).[3] Situated within a metanarrative of progress emerging out of the European enlightenment, this notion of the sublime invested electricity with 'the aura of divine force and utopian gift' and characterized it 'as the progenitor of a new era of social life' (Carey and Quirk, 1970: 396). Informed by an undercurrent of religiosity in the new world of America, this gift promised to solve the social and psychic problems and dislocations stemming from industrialization. It offered a combination that sought to fuse opposing poles: 'the desire for peace, harmony, and self-sufficiency with the wish for power, profit, and productivity' (Carey, 1989: 207). This faith ignored the fact that it was the wholesale embrace of technology that gave rise to the upheavals of urbanization and industrial mechanization in the first place.

Drawing on a range of literary and philosophical sources, Carey and Quirk show how potent economic and political forces were concealed by rhetoric that celebrated technological progress. For instance, the mass production of the electric light may have 'magically' illuminated the world in a way not previously experienced. But, once the novelty of the 'electrical sublime' subsided with the passing of time, an examination of the historical record shows the 'real beneficiaries' as the 'electric light and power companies' who produced and sold bulbs and ran the regional power grids

(Carey, 1989: 129). This is a timely lesson given the aggressive jostling for market position now occurring in relation to Big Data applications, services and markets.

In recent times, Vincent Mosco (2004) offers the best example of a thinker who integrates cultural myths and political economy in the analysis of computing technologies and the Internet. His notion of the digital sublime serves to conceptualize the Internet and Web as mythic spaces in which seductive tales about an information revolution, or a new age of politics and markets, are spawned and circulated. Following Carey's example, Mosco approaches the hydra-headed myths of technical rationality and linear progress as evidence of a 'nostalgia for the future' (p. 15) that will never be; reversing the temporal direction but not the logic of nostalgia – a yearning for a past that never was (Davis, 1979). However, the digital sublime is about more than the cultural construction of 'mythscapes' (Bell, 2003), as the ideologically charged stories told in its name create meaning and action in social and economic life, shaping the horizons of reality for citizens, corporate executives, industry analysts, journalists, investors, regulators, politicians and policymakers. Published a decade ago, Mosco's arguments have proven prescient. The Internet and World Wide Web have failed to fulfil the prophecies of a golden age of political freedom of expression and the democratization of knowledge. Instead, there is ample evidence of creeping corporate control of online media and the enclosure of intellectual property, surveillance of digital communications by the state, pervasive commodification of user data, and financial accumulation through the exploitation of unpaid user labour (Andrejevic, 2007, 2013; Curran et al., 2012; Gillespie, 2007; McChesney, 2013; Morozov, 2011).

The context and history outlined here are necessary to make sense of the hype associated with Big Data as well as the unfolding competition to profit from this enthusiasm. This article starts down this path, suggesting that modern sport forms a notable component in the 'webs of significance' (Geertz, 1973) that generate the popular appeal of Big Data. Indeed, for well over a century, sport has developed and refined rule-based systems to ascribe value and meaning to data generated by individual and collective human pursuits, emphasizing records, statistics, quantification and the application of technology to measure and maximize performance. These widely broadcast and publicized systems highlight that our social and cultural worlds 'are massively subject to statistical accounts' and that technologies of numbering and scientific measurement have long played a significant role in the cultures of modernity (Balnaves and Caputi, 2001: 1).

The world's richest data set

> Baseball offers perhaps the world's richest data set: pretty much everything that has happened on a major-league playing field in the past 140 years has been dutifully and accurately recorded, and hundreds of players play in the big leagues every year. (Silver, 2012: 80)

Big Data now has its own famous face: Nate Silver. His belief in the power of Big Data is on display in his best-selling book, *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't* (2012), in which he offers an open-ended commitment that ironically lacks the precision of his many statistically verifiable forecasts, 'Big Data *will* produce progress – eventually' (p. 12; original emphasis). As the book's impressive sales figures indicate, Silver's formidable reputation as a pundit is based on an ability to identify patterns of probability across a variety of statistical data, including voting records, financial markets, weather forecasts, gambling, games and, as the quotation above indicates, sport. Most notably, Silver upset many Republican Party adherents in the lead-up to the 2012 US presidential elections, using his blog on *The New York Times* website, FiveThirtyEight, to

correctly forecast the outcome of all 50 states that went to the polls and the re-election of Barack Obama. This performance followed the 2008 election when he predicted the voting outcome for 49 of the 50 states. Silver has since left *The New York Times* to join the Walt Disney-owned global media sport giant, *ESPN*, where he continues to write about politics, sport and other topics.

Silver came to prominence when he developed a player performance analysis tool for Major League Baseball (MLB), Player Empirical Comparison and Optimization Test Algorithm (PECOTA), in 2002. PECOTA published its first set of projections in 2003 after the owners of Baseball Prospectus purchased it from Silver in exchange for stock in their company. According to Silver (2012), his system performed a 'little better than the other commercial forecasting systems' (p. 88) and continued to be developed in the years that followed, extending its coverage to the analysis of minor league players. As Silver acknowledges openly, the value of this system in talent identification was a confronting challenge to an extensive network of 'old-school' talent scouts. These paid scouts and their networks of personal contacts felt their influence and employment under threat by software-driven statistical systems designed by Silver and others, having always made recruitment decisions about young players and minor league competitors on the basis of practiced observations and considered judgements. (It should be noted that Silver credits these scouts with considerable experience-based insight that is complemented by statistical assessment.) This conflict over the use value and application of computer-based 'sabermetrics'[4] in the world of professional baseball was then dramatized in the 2011 movie starring Brad Pitt, *Moneyball*. The 'Moneyball era' (Schumaker et al., 2010: 32) has also been ushered into another professional sport discussed by Silver, basketball.[5]

A reading of Silver's *The Signal and the Noise* reveals that his achievement in baseball was two-fold. First, he recognized that baseball had long been an example of Big Data. The game and its many leagues had been producing 'bigger and bigger data sets' (Mahrt and Scharkow, 2013: 22) as each season passed for well over a century. Silver's efforts were directed towards applying the processing power of networked computing systems, advanced software and adjustable algorithms to create an evolving analytical system that had the capacity to 'search, aggregate, and cross-reference' (boyd and Crawford, 2012: 663) baseball statistics as each game and season passed. This task involved the collation of a massive range of batter and pitcher statistics ('more than 10,000 player seasons') and physical attributes ('height and weight'). These sorts of statistics were then cross-referenced and analysed to assess probable performance over time (leading to the formulation of 'the aging curve'). Silver's system, which has enjoyed varying levels of success, enabled the direct comparison of players over time and assisted in the identification of rising and previously ignored talent.

His second and less recognized achievement was to take the statistical record of modern baseball – 'the world's richest data set' (p. 79) that had circulated freely in the public domain since its inception – to generate a privately controlled database that was sold for company stock. This commodification of the historical record proceeded by turning openly available baseball statistics into a searchable and manipulable database, thereby constituting a form of saleable intellectual property. As with other techno-entrepreneurs, Silver accumulated private wealth from the engineering and control of digital technologies and data (Gillespie, 2007). This achievement is rendered opaque by his ascending celebrity and the notion that he is a 'soothsayer' of Big Data, locating knowledge and truth in an ocean of statistics and digitized information that makes little sense to many pundits and citizens. If the healthy international sales of Silver's book are taken as a guide, the digital sublime is evident in the curiosity about his ability to apply specialist statistical and programming skills to locate 'the signal in the noise' produced by colossal amounts of data. As with other data analysts, however, Silver's apparent ability to predict the future probabilistically needs to be considered alongside the larger strategic game being played here; that is, framing the

perception and operation of emergent Big Data markets, services and technologies to ensure that enduring cultural legitimacy and financial profits are secured via machine-enabled analytical insights. This framing speaks to the broader advancement of commercial interests under the conditions of informational capitalism in which the successful fusion of technology and capital is central to success in the marketplace (Allen, 2013).

The opportunism of Nate Silver dovetails with the notion that we are living through a 'data rush' akin to the gold rushes of the 19th century (Mahrt and Scharkow, 2013). Money can be made by anyone canny, skilled and well resourced enough to identify the 'pure opportunities' on offer in the capture, control and organisation of digital data, which can then be marketed and sold (boyd and Crawford, 2012: 664). New books with exuberant titles such as *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Mayer-Schönberger and Cukier, 2013) are reflections of this commodifying impulse and its cultural catalyst, the digital sublime. The race is on to lockdown existing and emerging data sets given the possible insights they may contain if cross-referenced correctly and/or analysed by a new algorithm, which will be paid for handsomely by industry, government or private citizens with identified needs. As will be suggested, this trajectory connects with the development of modern sport, a setting in which informal leisure and amateur competition gave way slowly to a globalized system of elite sport characterized by the comprehensive quantification of performance, full-time professionalism and hyper-commodified media properties (Holt, 1989; Rowe, 2004, 2011).

Modern sport and the quantitative imperative

The emergence of modern sport occurred from the 17th century onwards alongside industrialization, urbanization and the spread of electrical technologies. Leisure-based pastimes, seasonal and religious rituals and informally organized games gave way to standardized rules and playing conditions, record-keeping and organized formal competitions occurring across an expanding number of regions and territories (Guttmann, 2004 [1978]; Holt, 1989). The 'sportization' of pastimes led to the export of rule-bound codes and competitions, including rugby, cricket, soccer and tennis, on an international scale with the advance of the British Empire, signalling a transformation in the organisation of time, space and culture that corresponds with the changing habitus of individuals under the conditions of modernity (Elias and Dunning, 1986: 22–34). In Allen Guttmann's (2004 [1978]) oft-cited Weberian analysis, this complicated history saw the peculiarly modern features of many popular sports emerge in the United States and elsewhere. The seven key characteristics he identifies are: secularism, equality, specialization, rationalization, bureaucracy, quantification and records. The last two characteristics are of particular relevance for my analysis because of the pervasive logic they introduce into the regulated and competitive movement of sporting bodies: 'Combine the impulse to quantification with the desire to win, to excel, to be the best – and the result is the concept of the record' (Guttmann, 2004 [1978]: 51).

A quantitative imperative infuses the pursuit of excellence in sport, with times, distances, heights, hits, throws, runs, goals and scores used to define and measure performance over time. As the experience of the 20th century shows, the combination of science and technology generates data deployed in the service of increasingly intricate systems that govern athletic bodies and minds in the quest for high performance, including systematic training regimes, new types of equipment, physiological surveillance, psychological management, and synthetic substances ingested to maximize performance or assist with recovery from injury. The interpenetration of machine-logic, mechanistic thinking and the human body (Rintala, 1995) accelerates with the advancement of

sports science, particularly during the Cold War when the pairing of quantification and the achievement of Olympic and World Records is bound to the official ideologies of the East and West (Hoberman, 1992). The sports science and training methods of the former Communist Bloc also find their way (in modified form) into countries such as Australia following the collapse of the Soviet Union. Newly available coaches are hired, carrying with them knowledge of material and computing technologies used for talent identification and effective training techniques (Hutchins, 2009).

Changes wrought by geopolitical manoeuvres and the state-funded scientization of sport (Hoberman, 1992) are then deepened and widened by the conduct of professional sport under the conditions of market-based globalization (Miller et al., 2001). The advent of personal computing from the 1970s onwards proved a boon for the field of sport statistics and the technologically enabled insights it offered across a number of domains, including baseball, basketball, soccer, cricket and even curling (Schumaker et al., 2010). Statistics became legitimised as a source of raw material for 'knowledge creation' throughout elite level sport (Schumaker et al., 2010: 29), utilized by coaches and sports scientists in a continual search for a winning edge over opponents. This search, in turn, produced a labour market for experts with high levels of 'database literacy' (Mukherjee, 2013) and skilled programmers able to design software and original analytical formulas. This market also contributed to an awareness among commentators, journalists and aficionados that statistical analysis is a source of knowledge and fuel for the 'sports chatter' that constructs the discursive rhythms of media sport cultures (Eco, 1986; Rowe, 2011).

This brief history helps to explain why ever more fine-grained measures of performance – or the data-isation of sport – continue to proliferate across major elite competitions. Software and hardware systems and wearable media devices monitor and analyse the biorhythms of athletes as well as track their movements in real time through Global Positioning System (GPS) technologies. These technologies are indicators of a pluralization of the boundaries between technology and humans, and science and culture (Beck and Lau, 2005; Latour, 1993). This pluralization is manifest in the quantitative interpenetration of sports results and records *and* sports science measurements *and* athlete performance data. Sitting beneath the surface level quantification of sports, results and records are bigger and denser data sets that are aggregated, cross-referenced and matched in an effort to produce better performances, new records and data-driven interactive entertainment for fans. The appeal of mining multiplying sets of sports data parallels the appeal of Big Data generally, holding out the possibility of innovative insights in a world where subjective human judgement is confronted by a bewildering amount of new information and content on a routine basis. Yet, as the next section outlines, what remains largely unexamined are the ways in which these possibilities are promoted by industry, the corporate actors operating in this area, and the inequalities connected to the collection and commodification of data.

Corporate players, data asymmetries and entertainment

Mechanisms for the generation of constant streams of data are prioritized by leagues, teams, media companies, and technology service providers. This trend is noted by the influential multinational consultancy firm, PricewaterhouseCoopers, in recent 'outlook reports' for the global sports market (2011) and North America (2013). These widely circulated industry reports emphasize the escalating investment in technologically 'enhanced' fan experiences by major 'teams, leagues, event organizers, and facility operators' (2013: 4). For example, 'social and media connectivity' and information technology infrastructure are leveraged by smartphone and tablet apps that 'inform,

connect and entertain' fans by offering promotions and rewards programmes. The ability to collect, organize and cross-reference data generated by promotional, social networking and locative media activities holds great promise as, 'Data analytics will continue to improve the sector's market intelligence, which should lead to higher conversion and retention rates' (2013: 4). The sponsors of global soccer brands such as AC Milan and Arsenal are also capitalising on a 'social networking base', using 'data mining to help them develop content that is relevant to each [media] platform and each consumer segment' (2011: 20). Given its role as a global consulting firm, PricewaterhouseCoopers is vulnerable to accusations of 'speaking up' the affordances of data mining and analytics. Nonetheless, as with Nate Silver, these efforts function to embed the prized association between data generation, new market insights and untapped business opportunities. This framing is of little comfort to those sports unable to generate large-scale data because they cannot afford generous promotions and reward programmes, scalable and reliable in-stadium distributed antenna systems (DAS) and Wi-Fi networks, or the expense of developing specialist social media and location aware apps. This situation also makes it unlikely that high profile sponsors and advertisers will invest significantly in sports below the top level if they are unable to offer substantial data mining opportunities that add value to brands.

The cumulative advantages flowing from the ability to generate, share and recombine data inform the ascending *value* that is assigned to it. The market for the provision of sports data services has grown rapidly in size and complexity over the past decade, particularly with the rise of broadband Internet and wireless communications, portable and mobile computing, digital camera capture and measurement technologies, and real-time networked user interactivity. In the United States, for instance, STATS[6] started as a non-profit entity in the 1980s and has since morphed into a multinational operation owned by the private equity firm, Vista Equity Partners. This corporation boasts formal service agreements to provide data to some of the richest leagues and sports in the United States and internationally, including the National Football League, MLB, the Professional Golfers' Association Tour, National Association for Stock Car Racing, Union of European Football Associations, Nippon Professional Baseball League and the Canadian Football League.[7] These contracts deliver 'personnel management, player evaluation and game preparation' systems to leagues and teams that promise 'objective analysis', 'operational efficiency' and an 'integrated data platform' (STATS LLC, 2013a). In the case of soccer, these systems use digital tracking technologies to report on player performance during a game in real time through the application of 'complex algorithms' that claim to analyse 'accumulated data streams and compile meaningful information and insights with incredible precision' (STATS LLC, 2013b). Of particular note is the announcement of STATS as the new 'official player tracking partner of the NBA [National Basketball Association]', adding to its status as the NBA's 'official data distribution partner' since 2006 (STATS LLC, 2013c). A system of six cameras and STATS SportVU 'propriety software' is used by each NBA team to 'provide a continuous stream of innovative statistics based around speed, distance, player separation, and ball possession for detailed and targeted analysis of players and teams', thereby creating a 'new vernacular of the NBA' (STATS LLC, 2013c). The promotional rhetoric of STATS illustrates the arrival of Big Data and its repressed other, the digital sublime, in the men's elite sport sector. Up-to-the-minute data contain the timeless qualities of objectivity, efficiency, insight and precision, but only for the right price in today's data marketplace.

Asymmetric control over the capacity to generate large-scale data and commodify the value generated by its analysis extends beyond North America. In the United Kingdom and Europe, significant market influence in the supply of sports data is exercised by Opta (recently acquired by the Perform Group), a company that claims to be 'the world's leading sports data company'

servicing '300 customers in 40 countries' (Opta, 2013). Working across a range of sports, a focus on soccer and rugby clubs and competitions is evident in their client list, in addition to an array of broadcasters, media outlets and betting operators.[8] Notably, the building of bigger and bigger data sets over time forms an explicit component of Opta's marketing pitch, 'As the live data goes into our databases, it adds to the most comprehensive archive of sports data on the planet' (Opta, 2013). This claim reiterates the notion that massive data is a gateway to marketplace success in an age of relentless digital media activity.

In the smaller market of Australia, the emergence of Champion Data parallels that of STATS and Opta. The success of this Melbourne-based company has run alongside the increased value assigned to player performance data by clubs, leagues, the news media and broadcasters, with the popular Australian Football League (AFL) leading the way in this regard. Having commenced operations in 1995, Champion Data's 'footprint' has expanded beyond the AFL to encompass cricket, the rugby codes and soccer. They also supply data to elite netball competitions in Australia and New Zealand, one of only a handful of women's sports featured prominently in the promotional materials of data providers.[9] The success of Champion Data and the importance assigned to broad ranging data collection and analysis in sport has seen the AFL purchase a 49% stake in this data provider, although not all of the AFL's 18 clubs are entirely happy. Gaps between the data-rich and the data-poor exist not only between sports but also between clubs playing in the same league. An indication of this imbalance is discernible in the AFL. According to a study of Champion Data completed by Jessie Wilkie (2011), a handful of clubs with significant budgets are considered 'top end clients', whilst less well-financed clubs struggle to afford all of Champion Data's analytical services and products. A club spokesperson for one AFL club states, 'Champion's [business] model is everything comes at a reasonable price. We can only have a relationship with the Championship guys to a point because the price just surpasses what we have [available to us]' (quoted in Wilkie, 2011: 39).

This comment by an employee of a nationally prominent club is instructive, especially as the poorest clubs in the AFL still report annual revenue from operating activities in excess of Australian $30 million. The nexus between financial resources and the cost of accessing sports data analytics is clear in this context, suggesting that those without any 'access to the database' (Andrejevic, 2013: 17) – lower level professional competitions, semi-professional clubs and pursuits that rely on athlete 'pay-for-play' funding structures – are locked out of the moneyed world inhabited by STATS, Opta and Champion Data.

The asymmetry described here matters because opportunities in the media sport marketplace are increasingly structured around *access* to reliable and comprehensive data. Companies supplying sports performance data to elite coaches, trainers and sports scientists repurpose sections of their data to create popular online consumer services and applications. Examples include fantasy sports, real-time match statistics and detailed infographics that appear on the websites and mobile apps of leagues, teams and commercial sports news outlets. These developments perpetuate a sharply stratified 'attention economy' (Lanham, 2006) in which leading sports and media companies use data and technology services to monopolize the time and energy of fans on multiple screens. Possessing the technical and human resources to produce expansive content, they bombard users with unremitting information 'about the same established band of professional sports and leagues that have long held popular attention' (Hutchins and Rowe, 2012: 69). This capacity enables experts and followers to know more than ever about dominant sports, making it even more difficult for lower profile sports to gain a sizable foothold in the media sport market. Moreover, as these consumer services spread, so too does the amount and types of user data available to those corporations that collect and control it.

The relationship between data supply and commercial enterprise is exemplified by fantasy sports, which have almost become a stand-alone industry with 'participants in the millions and a financial impact in the billions' (Billings and Ruihley, 2014: 1). Access to the data that enable the operation of fantasy leagues has been the subject of legal conflicts over the rightful control of athlete data (Evans, 2007–2008; Grady, 2007), and spurred an international market for data sets and software packages that allow sports and media organizations to set up fantasy competitions (Hutchins and Rowe, 2012). The transformation of sports data into a capital asset by those able to access, collect and sell it across multiple media platforms is reflected in the appearance of 'datatainment', a dubious buzzword that describes the 'evolution of data from intelligence, to information, to entertainment' (Strachan, 2012). Coined by Richard Ayers, a digital sports media consultant, this term connotes the refinement and repackaging of player performance data for fan consumption and enjoyment:

> For the football fan, data – specifically statistics – are the bread and butter of football-obsessed life-styles. Statistics such as average pass completion over a season or shots off-target are quickly becoming more common in mainstream media and pub arguments. So, by feeding this appetite with infographics, data visualisation web apps and raw access to databases – football can use data not just to inform, but to entertain. (Walsh, 2012a; Also see Walsh, 2012b)

A 'football-obsessed lifestyle' has become intertwined with statistics and different modes of data visualization. This shift is highlighted by the success of award-winning smartphone and tablet apps such as *FourFourTwo* magazine's 'Stats Zone', which allows fans to choose their selected statistical focus and share the outcomes across social media networks.[10] These are fortunate developments for a global football brand such as Manchester City that can afford to be serviced by two leading international data providers, Opta and Prozone,[11] and which reportedly employs 10 full-time data analysts to work with their first team alone (Siu, 2013a, 2013b).

It is worth noting that 'feeding the appetite' of fans for data presents unexpected challenges even for affluent sports clubs and franchises. Expensively acquired athlete performance data packaged for public consumption and fan discussion is a potential source of commercial and advertising revenue. But the open availability of this data is also a trigger for conflict between the digital media, marketing and financial objectives of a professional sports team, and the on-field performance aims of coaching staff and athletes. For example, openly available performance data may inadvertently hand opponents vital hints about the tactical tendencies and positional weaknesses of a team. Sean Callanan, a digital marketing expert who works with a number of Australasian and North American sports and teams, explains:

> These stats are growing all the time and for the digital department it's a content 'beast'. We're always trying to produce more content for the fans. Some of those stats are just cool . . . [for example] seeing heat maps of where certain players are always shooting from [on the court]. But the battle is that, as this information becomes more accessible, the less likely they are to share it from a team operations and coaches point of view. They [coaches] start treating this as the 'secret sauce' and don't want it to go out if the opposition gets hold of it. (Sports Geek Podcast, 2013a)

Tension around the use and availability of data also links to the rights of the athletes whose exertions produce it, adding to the databases controlled by corporations such as Opta and Champion Data. Inexpert analysis of decontextualized performance data can produce misleading conclusions that affect the reputation and value of an athlete. Even the most complex algorithms that cross reference multiple datapoints have difficulty in accounting for all existing contextual

factors, such as the game day mindset of an athlete, the quality of a coach and the (in)consistency of their instructions, changing roles within the team over the course of a game, conditions in different sections of a playing arena, hidden injuries and illness, and the attitudes of individual referees and unpredictable interactions between game officials. The demand among sports followers for so-called datatainment is not matched by a willingness to publicize the biases and limitations on which data measurements are based. In other words, the gloss of precision and objectivity affected by a series of infographics or cascading numerical tables is a questionable substitute for proficient qualitative interpretation that confronts the ineradicable subjectivities of individual and team performances over time.

Conclusion

The fascination with Big Data in the professional sports industry continues. The realization that market power can be exercised through the generation of data, control of the value assigned to it and access to data was on display at a major industry gathering, the Sports and Entertainment Alliance in Technology Conference. According to a conference steering committee member, Sean Callanan, 'Everyone was talking Big Data' over the course of this two-and-a-half day event in Kansas City in the United States (Sports Geek Podcast, 2013b). This focus was reflected in presentations and workshops discussing 'Big Data as a Business Strategy' and the challenges of how to 'Build Your Big Data Road Map'. The sense of possibility attached to these expositions is matched by a recent CBS television news report delivered by a seemingly astonished journalist about the spiralling demand for real-time sport statistics and the growth of companies such as STATS (CBS, 2013). As has been argued, the excitement attached to Big Data is evidence that the digital sublime continues to resonate in the network society, fetishizing the wonders of new technologies and promoting an optimistic belief in the unrivalled insights that can be produced by digital media and the data-isation of human endeavours. In the case of sport, figures such as Nate Silver and companies such as Opta, Champion Data, Prozone SAP, and Sportradar are profiting from this phenomenon, capitalizing on a demand for extensive data analytics in the professional sport and media industries. This demand is amplified by the continuing spread of networked digital communications and mobile devices. In drawing an historical analogy, the major operators in the sports data sector are arguably akin to the electric light and power companies of the industrial age (Carey, 1989). The novelty of the data technologies and services they provide will inevitably pass with the passage of time, but their current accrual of market power is set to have a long-lasting impact on the contours of data trade and use in the global sport market.

The inequalities created by the circumstances described in this article demand ongoing investigation and critique. The excitement surrounding Big Data and the rise of multinational sport data corporations conceals the widening inequalities between wealthy sports and those leagues and teams who struggle to achieve subsistence level sponsorship, media coverage and spectator attention. The 'democratizing' effects of digital media and networked communications are not apparent in the case of professional sport, particularly given the financial, human and technology resource deficits that presently exist. Already strong elite men's sports continue to become stronger in a multiplatform media environment where the pressure to produce and distribute content is constant, making it difficult for many women's and semi-professional sports to gain greater attention (Hutchins and Rowe, 2012). An emphasis on collecting bigger and bigger data sets only makes the situation more challenging for those sports and athletes that have little hope of collecting significant amounts of data, let alone affording the services of a global sport statistics company.

The distance between the data-rich and data-poor is also unlikely to close anytime soon given the prevailing dynamics of the informational, global economy. Successful commercial operators thrive as they access the constant opportunities presented by digitally networked markets and enterprises, leaving those left behind to struggle and fend for themselves (Castells, 2000: 187). This winner-takes-all ethos is germane to on-field competition in elite sport, a sphere of social activity that is helping to construct the cultural legitimacy and commodity value of data through the commodification of records, history and the bolstering of a lucrative global market.

Notes

1. Following the example of boyd and Crawford (2012), this term has been capitalized to make clear that it is the phenomenon under discussion. In addition to this, the employment of capitals signals that it is the social and cultural significance of Big Data that forms the primary focus of analysis, not the technical specifications of the complex hardware and software systems owned and operated by information technology and digital media corporations.
2. These include financial incentives for companies that provide 'high-tech industry' employment opportunities, tax breaks and the building of network infrastructure.
3. It is worth noting that Carey was a noteworthy critic of McLuhan, having greater sympathy for the arguments and approach of another seminal Canadian communications scholar (and economic historian), Harold Innis (see Carey, 1968, 2004).
4. As explained by the Society for American Baseball Research (2013), sabermetrics is the use of statistics to 'search for objective knowledge about baseball'.
5. *Moneyball* was a Columbia Pictures release based on a book by Michael Lewis (2003) about the Oakland Athletics and their general manager, Billy Beane. Also, ice hockey is another sport thought to be in the 'infancy of sabermetric research' (Schumaker et al., 2010: 12).
6. See www.stats.com
7. See http://www.stats.com/league_relationships.asp
8. See http://www.optasports.com/who-we-work-with.aspx
9. See http://www.championdata.com/index.php/sport/netball. Other examples include mentions of the Women's Tennis Association and Women's National Basketball Association on the website of STATS (http://www.stats.com/tennis.asp http://www.stats.com/basketball.asp), and the appearance of the FIFA Women's World Cup and the FA Women's League amidst a long list of men's competitions covered by Opta (http://www.optasports.com/media/124062/final_opta_coverage_copy.pdf).
10. See https://itunes.apple.com/au/app/fourfourtwo-soccer-stats-zone/id453744566?mt=8. The real-time data used by this app are supplied by Opta.
11. Recently acquired by STATS, Prozone sells performance analysis services through database and tracking technologies. Based in the United Kingdom, Prozone has regional offices in Europe, the United States, Asia, South America and South Africa. See http://www.prozonesports.com/index.html.

References

Allen M (2013) What was web 2.0? Versions as the dominant mode of internet history. *New Media & Society* 15(2): 260–275.

Andrejevic M (2007) *iSpy: Surveillance and Power in the Interactive Era*. Lawrence: University Press of Kansas.

Andrejevic M (2013) *Infoglut: How Too Much Information is Changing the Way We Think and Know*. New York: Routledge.

Balnaves M and Caputi P (2001) *Introduction to Quantitative Methods: An Investigative Approach*. London: Sage.

Barthes R (1973) *Mythologies*. London: Paladin.

Beck U and Lau C (2005) Second modernity as a research agenda: Theoretical and empirical explorations in the "meta-change" of modern society. *British Journal of Sociology* 56(4): 525–557.

Bell DSA (2003) Mythscapes: Memory, mythology, and national identity. *British Journal of Sociology* 54(1): 63–81.

Billings AC and Ruihley BJ (2014) *The Fantasy Sport Industry: Games Within Games*. New York: Routledge.

boyd d and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Carey JW (1968) Harold Adams Innis and Marshall McLuhan. In: Rosenthal R. (ed) *McLuhan: Pro and Con*. Baltimore: Penguin, pp. 270–308.

Carey JW (1989) *Communication as Culture: Essays on Media and Society*. New York: Routledge.

Carey JW (2004) Introduction to the Rowman & Littlefield edition. In: Innis HA. (ed) *Changing Concepts of Time*. Lanham: Rowman & Littlefield, pp. 7–20.

Carey JW and Quirk JJ (1970) The mythos of the electronic revolution. *The American Scholar* 39(3): 395–424.

Castells M (2000) *The Rise of the Network Society*. 2nd ed. Oxford: Blackwell Publishers.

CBS (2013) Demand for instant sports info fuelling growth of stats LLC. 5KPIX, 12 December. Available at: http://sanfrancisco.cbslocal.com/video/9622834-demand-for-instant-sports-info-fueling-growth-of-stats-llc/ (accessed 12 December 2013).

Communication & Sport (2013) Special issue: Women and the sports media 1(3).

Crawford K (2013) Think again: Big data. *Foreign Policy* 10 May. Available at: http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data#sthash.dmeEiuIC.dpbs (accessed May 2013).

Curran J, Fenton N and Freedman D (2012) *Misunderstanding the Internet*. London: Routledge.

Davis F (1979) *Yearning for Yesterday*. New York: The Free Press.

Denzin NK (1987) On semiotics and symbolic interaction. *Symbolic Interaction* 10(1): 1–19.

Dumbill E (2013) Big data is rocket fuel. *Big Data* 1(2): 71–72.

Eco U (1986) Sports Chatter. In *Travels in Hyperreality: Essays*, Weaver W (Trans.). New York: Harcourt Brace Jovanovich, pp. 159–166.

Elias N and Dunning E (1986) *Quest for Excitement: Sport and Leisure in the Civilizing Process*. Oxford: Blackwell.

Evans S (2007–2008) Whose stats are they anyway? Analyzing the battle between major league baseball and fantasy game sites. *Texas Review of Entertainment and Sports Law* 9(2): 335–351.

Gane N (2006) Speed up or slow down? Social theory in the information age. *Information, Communication & Society* 9(1): 20–38.

Geertz C (1973) *The Interpretation of Culture*. New York: Basic Books.

Gillespie T (2007) *Wired Shut: Copyright and the Shape of Digital Culture*. Cambridge: The MIT Press.

Grady J (2007) Fantasy stats case tests limits of intellectual property protection in the digital age. *Sport Marketing Quarterly* 16(4): 203–231.

Gregg M (2015) The gift that is not given. In: Boellstorff T and Maurer B (eds) *Data: Now Bigger and Better!* Chicago: Prickly Paradigm Press.

Guttmann A (2004 [1978]) *From Ritual to Record: The Nature of Modern Sports*. New York: Columbia University Press.

Hargreaves J (1994) *Sporting Females: Critical Issues in the History and Sociology of Women's Sport*. London: Routledge.

Hoberman J (1992) *Mortal Engines: The Science of Performance and the Dehumanization of Sport*. New York: The Free Press.

Holt R (1989) *Sport and the British: A Modern History*. Oxford: Oxford University Press.

Hunter M (2006) McLuhan's pendulum: Reading dialectics of technological distance. In: Vasudevan R, Sundaram R, Bagchi J, Narula M, Sengupta S, Lovink G, and Stikker M (eds) *Sarai Reader 3*: *Shaping Technologies*. Delhi and Amsterdam: Sarai, CSDS & The Waag Society for Old and New Media, pp. 144–156.

Hutchins B (2009) 26 January 1981. The Opening of the Australian Institute of Sport: The Government Takes Control of the National Pastime. In: Roberts D and Crotty M (eds) *Turning Points in Australian History*. Sydney: University of New South Wales Press, pp. 198–210.

Hutchins B and Rowe D (2012) *Sport Beyond Television: The Internet, Digital Media and the Rise of Networked Media Sport*. New York: Routledge.

Lanham RA (2006) *The Economics of Attention: Style and Substance in the Age of Information*. Chicago: University of Chicago Press.

Latour B (1993) *We Have Never Been Modern*. Cambridge: Harvard University Press.

Lewis M (2003) *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton & Co.

Lorentz A (2013) Big data, fast data, smart data. *Wired*, 17 April. Available at: http://www.wired.com/insights/2013/04/big-data-fast-data-smart-data/ (accessed July 2013).

Lycett M (2013) "Datafication": Making sense of (Big) data in a complex world. *European Journal of Information Systems* 22: 381–386.

McChesney R (2013) *Digital Disconnect: How Capitalism is Turning the Internet Against Democracy*. New York: The Free Press.

McLuhan M (1964) *Understanding Media: The Extensions of Man*. London: Routledge & Kegan Paul.

McMillan R (2013) The big data interview: Making sense of the new world order. *Wired*, 3 April. Available at: http://www.wired.com/wiredenterprise/2013/03/big-data/ (accessed July 2013).

Mahrt M and Scharkow M (2013) The value of big data in media research. *Journal of Broadcasting & Electronic Media* 57(1): 20–33.

Manovich L (2012) Trending: The promises and the challenges of big social data. In: Gold MK (ed) *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, pp. 460–475.

Maxwell R and Miller T (2012) *Greening the Media*. New York: Oxford University Press.

Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston and New York: Houghton Mifflin Harcourt.

Miller T (2012) *Blow Up the Humanities*. Philadelphia: Temple University Press.

Miller T, Lawrence G, McKay J, et al. (2001) *Globalization and Sport: Playing the World*. London: Sage.

Morozov E (2011) *The Net Delusion: The Dark Side of Internet Freedom*. New York: Public Affairs.

Mosco V (2004) *The Digital Sublime: Myth, Power, and Cyberspace*. Cambridge: The MIT Press.

Mouncey P (2012) Editorial: Wrestling with big data. *International Journal of Market Research* 54(4): 443–450.

Mukherjee R (2013) Interfacing data destinations and visualizations: A history of database literacy. *New Media & Society* 16: 110–128.

Opta (2013) About Opta. Available at: http://www.optasports.com/about/who-we-are/about-opta.aspx (accessed November 2013).

Pearlstein J (2013) Information revolution: Big data has arrived at an almost unimaginable scale. *Wired*, 16 April. Available at: http://www.wired.com/magazine/2013/04/bigdata/ (accessed July 2013).

PricewaterhouseCoopers (2011) *Changing the Game: Outlook for the Global Sports Market to 2015*. UK: PricewaterhouseCoopers LLP.

PricewaterhouseCoopers (2013) *At the Gate and Beyond: Outlook for the Sports Market in North America through 2017*. US: PricewaterhouseCoopers LLP.

Rintala J (1995) Sport and technology: Human questions in a world of machines. *Journal of Sport and Social Issues* 19(1): 62–75.

Rowe D (2004) *Sport, Culture and the Media: The Unruly Trinity*. 2nd ed. Berkshire: Open University Press.

Rowe D (2011) *Global Media Sport: Flows, Forms and Futures*. London: Bloomsbury Academic.

Schumaker RP, Solieman OK and Chen H (2010) *Sports Data Mining*. New York: Springer.

Silver N (2012) *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't*. New York: The Penguin Press.

Siu V (2013a) Proliferation of data-driven analysis in soccer (Part 1: The Scouts). *SportTechie*, 20 November. Available at: http://www.sporttechie.com/2013/11/20/the-proliferation-of-data-driven-analysis-in-soccer-part-one-the-scouts/ (accessed December 2013).

Siu V (2013b) 'Proliferation of Data-Driven Analysis in Soccer (Part 2: The Coaches)', *SportTechie*, 20 November. Available at: http://www.sporttechie.com/2013/11/22/the-proliferation-of-data-driven-analysis-in-soccer-part-two-the-coaches/ (accessed December 2013).

Society for American Baseball Research (2013) A guide to sabermetric research. Available at: http://sabr.org/sabermetrics (accessed October 2013).

Sports Geek Podcast (2013a) SGP 012: Tod caflisch from red wings, trendsmap and welcome Mr Cricket', 28 July. Available at: http://sportsgeekhq.com/podcast/sgp-012-tod-caflisch-from-red-wings-trendsmap-and-welcome-mr-cricket/ (accessed August 2013).

Sports Geek Podcast (2013b) SGP 014: #SEAT2013 mega interview podcast', 21 August. Available at: http://sportsgeekhq.com/podcast/sgp-014-seat2013-mega-interview-podcast/ (accessed August 2013).

STATS LLC (2013a) ICE basketball. Available at: http://www.stats.com/ice_basketball.asp (accessed November 2013).

STATS LLC (2013b) Data – Soccer/football. Available at: http://www.stats.com/sportvu/football_data.asp (accessed November 2013).

STATS LLC (2013c) NBA expands partnership with STATS LLC to unveil player tracking technology for All 30 NBA teams. Press Release 5 September. Available at: http://www.stats.com/pr_090513.asp (accessed November 2013).

Strachan D (2012) Datatainment: How data can engage sports fans. *UKSN: Social Media & Sport*, 9 July. Available at: http://www.theuksportsnetwork.com/datatainment-how-data-can-engage-sports-fans (accessed November 2012).

Walsh S (2012a) Manchester city datatainment – player data goes public. *Digital-Football.com*, 16 August. Available at: http://digital-football.com/featured/manchester-city-datatainment-player-data-stats-go-on-the-website/ (accessed November 2012).

Walsh S (2012b) "Datatainment" the new trend for #Euro2012 & sports social media. *Digital-Football.com*, 19 June. Available at: http://digital-football.com/featured/datatainment-the-new-trend-for-euro-2012-sports-social-media/ (accessed November 2012).

Wilkie J (2011) *Champion data meets AFL: Exploring their media driven relationship*. Unpublished Honours Thesis, Deakin University.

Author biography

Brett Hutchins is an Australian Research Council (ARC) Future Fellow and Associate Professor in Media Studies in the School of Media, Film & Journalism at Monash University in Melbourne, Australia. His recent co-authored and edited books include *Digital Media Sport: Technology, Power and Culture in the Network Society* (with David Rowe; Routledge 2013), *Sport Beyond Television: The Internet, Digital Media and the Rise of Networked Media Sport* (with David Rowe; Routledge 2012), and *Environmental Conflict and the Media* (with Libby Lester; Peter Lang 2013).