Cultural Studies Review
volume 20 number 1 March 2014
http://epress.lib.uts.edu.au/journals/index.php/csrj/index
pp. 121–49

Data Motility

The Materiality of Big Social Data

MARK COTÉ
VICTORIA UNIVERSITY

One invaluable source for reading the code of the 'digital human' is mainstream business media. Ever since enterprise computing spread to desktops in the 1980s, the Californian ideology has imbued neoliberalism with dreams of ever-more profitable information technology, filling pages from the *Wall Street Journal* to *Wired*.[1] Now, with the rise of mobility, the explosion of data and the proliferation of platforms and apps, such appraisals continue to be breathlessly dispensed. Sometimes, however, there is a critical revelation in the assessments of profitability. Consider this trenchant maxim for understanding social media and big data recently offered by Tim Worstall, a fellow from the Adam Smith Institute: 'It's an old adage that if something is free it must be you that is the thing being sold.'[2]

I find this statement richly resonant. First, intentionally or otherwise, Worstall encapsulates a radical critique of the conflation of media production and consumption that stretches from Dallas Smythe's 'audience commodity' to Maurizio Lazzarato's 'immaterial labour'.[3] Second, and more to the point here, it stands as an affirmation of Foucault's 1975 methodological imperative to look beyond the 'great texts' to the archives of everyday life when looking for the effective discourse of power, as the bourgeoisie said 'precisely what it was doing, what it was going to do, and why … It stated perfectly what it wanted'.[4] I cite Foucault here not merely to

ISSN 1837-8692

bolster a blog post from *Forbes.com*; rather, it signals the theoretical paradigm fundamental to my analysis. The aforementioned quote marks the first time Foucault described his largely overlooked but vital concept: the *dispositif*. He developed it to move beyond the myriad limitations of a discursive analysis of power toward a more heterogeneous ensemble which includes the non-discursive. Taking the example of Foucault's disciplinary *dispositif,* it includes i) the discursive regulations of juridical processes, and ii) the non-discursive materiality of institutions like prisons and the panopticon. One might be tempted to say this marks an incipient 'new materialism' that has largely gone unappreciated in the later Foucauldian analysis of power.

In this article I will present such a new materialist interpretation and will use the *dispositif* as its conceptual frame. I present the *dispositif* as positioned on the following theoretical continuum. We can start with Deleuze, who considered the *dispositif* a conceptual friend and saw it inextricably intertwined with his notion of the assemblage.[5] In turn, the assemblage—*agencement* in French—is *the* cohering concept in actor–network theory (ANT), which expanded notions of agency to include nonhuman elements, 'prostheses, tools, equipment, technical devices, algorithms, etc'.[6] Understanding agency as distributed across human–nonhuman assemblages is a hallmark of new materialism. Such assemblages, as deftly outlined by Dolphijin and van der Tuin, are critical to the development of materialist feminist theory (by, for example, Grosz, Braidotti and Barad, among others) which proffer a nonrepresentational theory of power.[7] The key here is the affordance of a dynamic role of desire, which, for materialist feminist theory, could account for a non-essentialist understanding of sexual differing, as opposed to sexual difference. Dolphijin and van der Tuin cite this specific instance to highlight a more general importance for new materialism, underlining it with a key passage from Deleuze: 'it is not the dispositifs of power that assemble [agenceraient], nor would they be constitutive; it is rather the agencements of desire [desiring-assemblages] that would spread throughout the formations of power following one of their dimensions'.[8] Such a conceptual orientation makes visible the diffusion of agency and desire/intentionality across a *dispositif*. If applied to the mediated environment of the digital human, the *dispositif* brings into focus the dynamic tension between

communicative creativity and its capture, marking out the interplay between sociality and capital therein.

Here I present the *dispositif* of 'data motility' for such a new materialist analysis of the digital human, the discursive and non-discursive assemblage of the 'you' being sold. Motility denotes how the data you generate increasingly moves autonomously of your control. The assemblage comprising this dispositif, however, must be critically unpacked, lest it remain an analytical 'black box'. This may be desirable for corporate interests dealing in big data, but it does little for an informed understanding of life in the age of big social data (BSD). A sustained and rigorous analysis is beyond the scope of this article, and I am undertaking such a larger project elsewhere. As noted, my intention here is to introduce the *dispositif* as a conceptual frame for the study of BSD and the digital human. I will, then, identify a few non-discursive or material elements comprising that assemblage, namely the kind of data which makes up BSD, and the weight and structure of the cloud through which it moves.[9] Finally, I identify for further study how the deeply recursive materiality impacts upon the life, labour and debt of the digital human under BSD.

Throughout, I am sensitive to what I see as the 'desiring-assemblage' of motility, the movement of the BSD we produce. For, indeed, if it is our digital selves that are being sold, I am not suggesting this happens simply in the instrumental service of digital capital. Therein lies one of the great benefits of the *dispositif* as a critical methodology: its assemblage coheres in a dynamic of tension and struggle, without a singular, instrumental driving logic or a sedimented hierarchy. Practically, this means that the 'you' being sold—the social data we all generate—is motile; that is, it flows from us, through our myriad personal technological artefacts and the material intricacies of the cloud initially as an expression of sociality. Yet its movement is not directed by us, and is almost wholly autonomous of our control. Indeed, the data we generate increasingly is moving at the behest of capital and the state. To put a finer analytic point on this, we might make a critical distinction between motile and mobile. Thus we can consider the contained movement of data that primarily augments the profitable growth of the business of BSD and new forms of digital state surveillance as data *mobility.* Yet there is a glitch inherent in the movement of data as the material environment of the cloud results in the seemingly self-directed movement of data itself. I read this both as a metaphor of the inherent

sociality of data and as a practical example that, invariably, all data enclosures leak in all directions. As such, data *motility* signals a possible route for the progressive becoming of a new data commons. It is my contention that the *dispositif* of data motility—along with its counterpoint mobility—can help us understand our collective stakes in the kinds of contestation inherent in data motility.

—WHAT IS A DISPOSITIF?

One of the most compelling reasons for using the *dispositif* to conceptually frame the life of the digital human under BSD is the importance it accords both to materiality and to thinking in terms of a complex, heterogeneous ensemble. It is important to note the *dispositif* marks an overtly politicised shift by Foucault, away from the structuralism and hermeneutics that defined his work through *Archaeology of Knowledge*. In his engaged political projects of the early 1970s like the *Groupe d'Information sur les Prisons*, and in his writings and interviews, Foucault acknowledged the methodological malaise that arose from a solely discursive focus as well as theorising power as domination. In 1975 Foucault acknowledged, 'I was in a dead end. Now, what I would like to do is to try to show that what I call the dispositif is something much more general than the episteme.'[10]

Foucault's turn to the *dispositif* began in *Discipline and Punish*, and became overt in *History of Sexuality* vol. 1, given that the organising concept of the latter was the *dispositif de sexualité*. It is unsurprising that this was overlooked by most English-language interlocutors because *dispositif* was inconsistently translated—as apparatus or mechanism or deployment—which obfuscated its conceptual importance. What should be clear, however, is the decisive move beyond the symbolic and representation. When describing his approach in *Discipline and Punish*, Foucault noted his analysis now included a 'thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic'.[11] This is the point where Foucault fully nuances power as symbolic and material, as relational, as microphysical, as circulating in networked formations, and not as simply a repressive force which says 'no'. Hence the importance of the *dispositif de sexualité,* not as a repressive Victorian ideology, but as a discursive-non-discursive matrix through which a normative (and, consequently, 'abnormal')

sexuality becomes visible and articulable. This indicates how the dispositif is to be understood as an analysis of power. In this reconceptualisation, Foucault is rejecting power as that which is centrally located, in the mode of production or in the state; nor is it a fungible commodity possessed by individual subjects and wielded like a club. Instead, it is expressed in heterogeneous ensembles, in complex assemblages of the discursive and non-discursive, of power and knowledge, and through which processes of subjectification or individuation unfolds.

I want to make two more quick points before identifying the non-discursive elements of the *dispositif* of data motility. In one of his first references to a concept that would retain sustained interest, Foucault described biopower as a *dispositif*: 'The biological traits of a population became relevant elements for economic management, and it is thus necessary to organise around them a dispositif which assures not only their subjugation, but the constant increase of their utility.'[12] I cite this because the productive dynamic of this *dispositif* is continued in that of data motility, wherein our quotidian actions have become discrete quanta, visible through their digital traces, and constantly subject to circulation in ways that increase their 'utility'. What is missing from biopower—and Foucauldian *dispositifs* in general—is recognition of the intimate relation between the body and mediating technology. This aporia is addressed by the dynamic presence of data motility.

The other point is in the polyvalent nature of power expressed above. Some readers may be thinking that biopower, in fact, was used by Foucault to indicate a rather repressive force, and they would be right. Lazzarato again helps here, noting that we must distinguish biopower from biopolitics. Specifically, biopower is a *dispositif* of control and domination, whereas biopolitics is a domain of creativity and resistance.[13] It is in following this model that I distinguish the contained and constituted flow of data mobility from the deterritorialising and nomadic flow of data motility. The *dispositif*, then, is not underwriting a utopian analysis, seeking out only lines of optimism in these heterogeneous ensembles. Rather, it is riven by struggle, and contained within the assemblage of a *dispositif* is both an analytic and a diagnostic of power, enabling critique of what we are and identifying what we might become. As Deleuze notes, the analytic of power examines 'what we are (what we are already no longer)' while the diagnostic considers 'what we are in the process of becoming'.[14] The *dispositif*, then, identifies the ways we are amid relations of

domination, but not in a manner that leaves us permanently trapped. What is most *a propos* to the study of BSD here is the role of the archive.[15] On the one hand the archive is the sedimented part, like the nineteenth-century prison studied by Foucault, the realm of the analytic of power. Yet Foucault equally identifies the archive as that 'which is at the same time close to us, but different from our present; it is the border of the time which surrounds our present'.[16] It is a liminal zone between what is sedimented and becoming. The archive—which is certainly one compelling way that digital traces of BSD can be framed—is a key fulcrum point in the strategic value of the *dispositif*. The motility—the movement—of that data can *both* reinscribe and reproduce relations and patterns of domination, *and* provide the material for creative resistance and becoming. In this sense, BSD is a site of struggle, and the manner in which our data circulates therein is of vital importance.

—THE MATERIALITY OF THE CLOUD

> Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.[17]

The *dispositif* of data motility makes visible and enunciable the movement and machinations of BSD. A focus on its non-discursive or material qualities brings into focus the nearly inconceivable volume of BSD, and the velocity with which it is captured and grows. Size and speed are key factors in its valorisation, and while economic value drives capital to maniacally increase the capture and analysis of BSD, it is the pursuit of social and cultural value that drives its generation. By foregrounding these tensions, I will try to make sense of data motility first by examining the kind of data that comprises BSD. One could say this means delineating BSD via the materiality of its discursivity which uneasily coexists in forms both machine readable and human readable. I then examine the architectural form in which it is stored and through which it gains motility. In short, I introduce both the kinds of data and the databases through which motility transpires. What follows, then, is an introduction to the materiality of the kind of BSD produced, and the structure of its archives.

Even a cursory quantification of the BSD produced by the digital human is challenging, given its rapid growth. Only ten years ago, humanity collectively generated about five exabytes of data per year. For clarification, one exabyte is the equivalent of one million terabytes. In 2012, we generated 2.7 zettabytes (2,700 exabytes), and it is predicted that by 2020 we will reach 40 zettabytes annually. That is an increase of 8000 times over in less than two decades.[18] I should distinguish what I am calling BSD from the broader category of 'big data'. The latter is more inclusive, entailing sensor data from industrial and domestic networks, radio-frequency identification (RFID) tags and 'The Internet of Things', financial markets, and big science projects, among others. While all this data contributes significantly to the quantification of the world in which we live, the focus here is on the social data generated by the digital human. BSD, then, comes from the mediated communicative practices of our everyday lives, whenever we go online, use our smartphone, use an app or make a purchase. Consider just three of the most popular sites. Google, back in 2008, the most recent available statistics, was processing 20 petabytes per day. In 2012 Facebook users were sharing four billion pieces of content per day, three billion things were 'liked' and three hundred million photos uploaded. Overall, Facebook's one billion users generate five hundred terabytes of social data every single day. Twitter sees nearly two hundred million tweets per day. Finally, there are now five billion people calling, texting, tweeting, browsing, posting and generating content on their phones.[19] Schematically, then, we see the smart phone as a key new vector of mobile communicative sociality, and that user-generated content primarily transpires on proprietary platforms. What might we glean from the materiality of that data?

Before the rise of social media and mobile computational power, much of the information that was stored digitally was structured data. This is data input into fixed fields, like columns or rows, each of which is clearly defined, as are their relations to one another. Spreadsheets are a quintessential form of this, termed a relational database. The information—or at least each discrete quantum—is simple and uncomplicated, insofar as it is relevant only to its field. Think of demographic information, like your address or date of birth, put into a spreadsheet, as singular forms of structured data. This format means structured data can be accessed and queried by, for example, structured query language (SQL) because of its clearly

identifiable and pre-defined schema. Prominent kinds of structured, relational databases include retail transaction data, financial market data, and industry, medical or pharmaceutical research data. Several things should be emphasised about structured data and its operating environment, called a relational database management system (RDBMS). First, it was long the preferred form of data, especially by corporate IT, because its highly predictable structure allows it to be efficiently processed. This efficiency results from the data being structured not for human but machine readability. Such structured data often would be input by a data entry clerk into a bespoke and costly environment like that provided by Oracle. I make this point to emphasise that structured data is typically instrumental, highly focused and subject to a pre-defined data model, always intended for efficient processing. While its content may very well represent elements of everyday life, it would not typically be produced, as data, through quotidian communicative sociality. In other words, traditional structured data is more typically composed for a functional purpose, and from inception is structured in a manner that machines like. Structured data, then, is structured vis-à-vis the symbolic realm of computation, of codes, programs and algorithms.

Humans, on the other hand, largely communicate in the symbolic realm of cultural meaning. We do so regardless of the specific non-human or technological elements with which we are assembled—although it must be noted that the historico-medium specificity has profound epistemological and ontological effects. The materiality of our augmented communicability, manifested in BSD, attests to this key historical difference and it illuminates just how data motility transpires. BSD is *not* new insofar as it emanates from the kind of communicative sociality that has always been endemic to the human condition. What is new and why it is of such importance to media theory are the particularities of its technological mediation. The newness of BSD, then, first comes in the form of the quintillions of raw data points being generated every day, which are captured and contained primarily by capital and the state, and proprietarily available for potentially never-ending future analysis. What is also new is that even though it is generated through personal computational devices, it is not in the efficient, machine-readable form of structured data. There is a longstanding rule of thumb that more than 80 per cent of the data we generate is, in terms of computer processing, unstructured.[20]

To clarify, unstructured data is not produced in pre-defined fixed fields, residing in relational databases. At its point of generation, user-generated social data does not conform to a pre-defined schema or data model for processing even though it is generated in the structured space of its platform. Rather, it is generated as informational or affective symbolic content, the result of spontaneous, contingent, free-form communicative sociality. BSD is unstructured data because it comprises the traces of the cultural life of the digital human. These are the textual objects that you generate in a blog, social media, a search, a message or an app; they are also the bitmap objects, the images, photos and videos that you send, post, like or tag.

Some debate the validity of the term 'unstructured' because if data were truly unstructured it would be unreadable gibberish in any format, by humans and machines. Further, a strong claim can be made that data is always 'structured' when entered into any digital realm. Every website, platform, or application is always comprised of a template created by software and information architects. The insistence on the fundamentally 'structured' nature of data is a shibboleth among proponents of software studies, ranging from Galloway's 'proctological wrappers' to Mika's application of the semantic web to social networking to Gehl's 'real software abstractions'.[21] These important contributions, however, can unintentionally obscure key changes in the material makeup of BSD, especially regarding its computational infrastructure. This distinction is most clearly exemplified by contrasting the newer Hadoop cluster to the older RDBMS environment. There is great analytical value in retaining the working distinction made by most computer scientists between structured, unstructured and semi-structured data. This is a distinction upon which I will build to better enact a materialist analysis of data technologies; that is, to outline what is new about the big social database as a medium. Such a distinction helps illustrate the new paradigm of computational power—social, political and economic—that emerges in the big data-crunching environment of Hadoop.

To risk further complicating matters, there is a third category: semi-structured data. This typically refers to things like XML (extensible markup language) and its simpler Java-script counterpart JSON, which encodes web documents in a manner both human- and machine-readable. These are basic tags and markers that give some structure to documents and facilitate information exchange. This is extremely

important for downstream processing and aggregation, the very interchange of heterogeneous data sources that is integral to data motility. These distinctions then, regardless of their disputed status, help to delineate the important material differences marking BSD. The challenges that these different forms of data create for their efficacious processing are important for my critical analysis as they help circumscribe the very conditions of motility.

For the moment, let's put aside the challenges the average digital human faces in translating and comprehending the interplay of the different forms of data she produces. Instead, let's consider the challenges faced by big data companies and social media giants like Google and Facebook in translating the unstructured data that humans produce into structured data that can be processed at speed and on a vast scale. For BSD makes particular infrastructural demands. One way to understand this paradigmatic shift is to trace a material link in the explosion in BSD back to a desktop-bound curiosity, the University of California-Berkeley's SETI@home (Search for Extra-Terrestrial Intelligence). This distributed computing project was one of the first examples of internet-scale applications, established back in 1997. Within a few years distributed computing took a pronounced cultural turn: Napster emerged, and its peer-to-peer file sharing successors—be it the bit-torrent protocol of Pirate Bay or the file-hosting service of Megaupload—made the widespread exchange of data a prominent new mediated practice. This was further intensified by the emergence of Facebook, YouTube, Twitter, social gaming like Farmville and e-commerce like eBay, applications and platforms that all scale to global reach and demand. When we add to that the rise of mobile devices and ubiquitous connectivity, the environment for the quotidian generation of BSD, be it structured, unstructured or semi-structured becomes clearer.

The internet-scale applications of social media via mobile devices alone created data footprints that were ill-fitted for traditional RDBMS, not just in terms of volume, but also because of the need to integrate different kinds of data from different sources. In short order, there emerged an urgent need to be able to access and aggregate multiple data sets on a vast scale, which required changes in computer architecture and network capacity. I should add that Foucault conceived *dispositifs* as assemblages which cohere in response to an *urgent need*. He writes that the *dispositif* is 'a formation which has as its major function at a given historical

moment that of responding to an urgent need. Thus the dispositif has a dominant strategic function'.[22] It is worth recalling again his *dispositif* of biopower and the urgent need to which it responds: 'the assimilation of a floating population found to be burdensome for an essentially mercantilist economy: there was a strategic imperative acting here as the matrix for a dispositif'.[23] But just as with the contradictions and tensions between biopower's *dispositif* of control and domination, and the creativity and resistance of that of biopolitics, I will suggest that the urgent needs of Google et al. differ considerably from those of the digital human.

Google is at the architectural heart of the rise of this data intensive computing environment. As its search engine became the near *de facto* mode of seeking internet-based content, the operational demands placed on its page-rank algorithm intensified. Already by the early 2000s, Google was struggling with its core business: the daily indexing of the entire web necessary for optimising the aforementioned algorithm. In order to cope, it radically reconfigured its approach, shifting to parallel processing distributed across vast networks. A series of papers in 2003 and 2004 by Google engineers helped rearticulate that company's hardware and software and, in the process, map out the environment in which BSD would flourish. In short, Google established a new paradigm for processing big data. They outlined a platform on which could be built the massive indexes from the internet for real-time analysis by extrapolating from the fundamentals of distributed computing. Think back to the SETI@home project which ingeniously managed a computational task that would have been prohibitively expensive from a central site: analysing the universe for signs of extraterrestrial life. By taking vast observational data from the Arecibo radio telescope, breaking it down into small chunks, and then having it analysed by home desktop computers, it proved the practical value of distributed processing. Similarly, Google needed to process the search requests that were scaling up at a rate similar to that of data in general: from 9800 requests daily in 1999 to 60 million in 2000 to 200 million in 2004 to 4.7 billion in 2011. By developing the Google File System and MapReduce, which are the core of the Google app engine, it addressed this urgent need to 'parallelize the computation, distribute the data, and handle failures'.[24]

The Google File System is a proprietary scalable distributed file system designed to run on inexpensive commodity hardware and to be highly fault tolerant

and able to process massive and expanding amounts of data.[25] MapReduce establishes the computational paradigm for handling the processing and generation of Google's large data sets, comprised of raw data gathered from web crawling, web request logs, derived data summarising search queries, pages crawled and the graph structure of web documents. The paradigmatic breakthrough of MapReduce is in making practical the clustering of large numbers of commodity PCs for automatic parallel and distributed computation on a large scale.[26] So it is in Google's proprietary environment that the new paradigm in which BSD would flourish was established. Just as with SETI@home, massive data calculations are broken into small chunks across many computers and, when completed, are reassembled into a single dataset. This is the basic design behind Google's scores of proprietary, warehouse-sized computing facilities which operate like one giant mainframe.

Because Google published key papers detailing their file system and MapReduce—albeit keeping their code a proprietary secret—others were able to develop the basic structure of the file system and processing. Hadoop, housed under the not-for-profit Apache Software Foundation, developed an open source implementation of Google File System and MapReduce. While Hadoop was built and is maintained by a global community of participants, there are countless for-profit organisations that run the framework for their own proprietary large distributed computation platforms. Hadoop and these related companies provide the software and data processing systems that enable the distributed computing that transpires on 'the cloud'. Reckoning the competing definitions of the amorphous computing cloud recalls Joseph Conrad in *Lord Jim*: 'the simplest impossibility in the world; as, for instance, the exact description of the form of a cloud'.[27] Yet this brief material overview reveals several key elements that can be described, and which detail this paradigm shift as it relates to data motility.

What has changed, and is important about the Hadoop cloud as a computing environment for BSD is i) the scalability of computing, ii) the new economics of storing data, iii) the ability to continuously question raw data, and iv) the emergence of raw data as a heterogeneous source for potentially endless aggregation. Amr Awadallah, a former Yahoo engineer and chief technical officer of Cloudera, a Hadoop-based private company, has cogently outlined these elements. The first depends upon the aforementioned distributed model. What must be stressed is the

computational power that comes from cluster architecture; that is, when a large number of computers are networked to run as if they were a single system. A simple example demonstrates the exponential power of the cluster. Say the single hard disk of a commodity PC can process one gigabyte per second, and one server holds twelve disks, and a rack holds twenty servers; that is already a processing speed 240 times faster than the single PC. Now the average cluster holds six racks, making it 1440 times the processing speed. If you move into the realm of large clusters, which big data and social media companies would typically deploy, you are suddenly processing 4.8 terabytes per second, some 48,000 times faster than a single PC. In practical terms, a large cluster can process in one second what would take thirteen hours on a single PC.[28] In the simplest terms, the larger you scale up, the faster your processing speed. The computational power of the cluster architecture is a potential resource awaiting more widespread and non-corporate deployment, and could enable a more inclusive and distributed community-based access to BSD.

In addition to upwardly scalable processing speed is a new economics of storage costs. In 1980, it cost US$193,000 to store one gigabyte of data; that would make one of today's sixteen gigabyte flash keys worth just over $3 million. By 1989 it was $36,000 per gigabyte, down to $43 in 1999 and to about six cents today.[29] Whereas an older corporate computing paradigm operated on return on investment (ROI) as a function of the cost of storing that byte, now it is return on byte (ROB), and given the relative pittance for storage, the basic question is how much value is created from the data you collect? This key change in the materiality of data storage carries a straightforward new imperative: collect more data. Further, as Awadallah notes, this new economics of 'keeping the data alive' also underpins the third fundamental shift of retaining the 'original raw event data'.[30] The cluster architecture, then, enables a new economy which maximises both the storage capacity and processing speed of data, and retains data in its original high-fidelity, unadulterated form for continuous future queries. In other words, structured, unstructured and semi-structured data are always available in their original form. In the traditional RDBMS, raw data is moved from the storage-only to the computational grid, where it is converted into the required structured form for database processing. But it is extremely expensive to reverse the process and retrieve the original data for further processing. The Hadoop environment, however,

makes no such distinction between storage and computation in its cluster architecture. Indeed, it requires no pre-defined schema or structure for its data, which can be taken from smart phones, RFIDs or the internet and dropped into the Hadoop cloud. This flexibility greatly diminishes the former challenge of processing structured, unstructured and semi-structured data in the same environment.[31]

Quite to the contrary, the heterogeneity of data becomes a potential virtue, insofar as it vastly widens the conditions of processing possibilities. With the imperative to collect more data built in to the material structure of a Hadoop environment, the ROB ratio becomes extremely attractive. That is because in straightforward economic terms, the original raw event data is now forever. The Hadoop structured cloud affords the cost-effective ability to store all forms of data now and process it later, and then process it again and again. The implications for BSD are significant. It means that data need no longer be considered a monolithic block for pre-determined processing, as was the case with most RDBMSs. It means an end to what is known as 'data exhaust'—the myriad forms of data which are stored temporarily and then deleted—will increasingly be a thing of the past. The archives of the digital human, as such, will continue to grow apace. The breadth and depth of the totality of BSD becomes in practice discrete data points wherein the possibilities for aggregation and analysis depends only on the imagination of those querying the data. In this sense, surely it is critical that this questioning not be left exclusively in the realm of marketers. A very brief look suggests an avalanche of ideas, all designed primarily to increase our efficacy, that is our profitability as consumers.

The material elements comprising data motility are highly conducive to the needs of capital. The 'powered by' page on the Hadoop Wiki reads like a who's who of social media, e-commerce, advertising, marketing and broadly defined BSD-related companies.[32] Yahoo runs Hadoop with over forty thousand nodes, including a single 4500-node cluster. eBay runs it for search optimisation and research; Last.fm and Spotify for data aggregation, reporting and analysis. Netflix also uses Hadoop to process the vast user-data it gathers from streaming programming, which it uses to integrate consumption even more deeply with production. Facebook runs the world's largest Hadoop cluster, about one hundred petabytes and capable of ingesting five hundred terabytes of new data every day.[33] Future research is

necessary to comprehensively outline corporate Hadoop users and the specific forms of data analysis they perform. Here, I simply want to isolate a telling element of Facebook's BSD infrastructure. Again, following Foucault's imperative, I turn to the business press and quote Jay Parikh, Facebook's vice-president of infrastructure engineering:

> We also use a version of Hadoop and Hive to run the business, including a lot of our analytics around optimising our products, generating reports for our third-party developers, who need to know how their applications are running on the site, and generating reports for advertisers, who need to know how their campaigns are doing. All of those analytics are driven off of Hadoop, HDFS, Hive and interfaces that we've developed for developers, internal data scientists, product managers and external advertisers.[34]

What Parikh highlights—optimising reports, generating app reports and reports for advertisers—are core practices of BSD analytics. The material infrastructure and practices we have been outlining are a necessary precondition for BSD analytics, be it as data mining, sentiment analysis, or predictive analysis. These new core practices are extensions and intensifications of the kinds of surveillance strategies of data exploitation so comprehensively outlined by Andrejevic and Fuchs.[35] While such data capture is manifest, the heterogeneity of the *dispositif* demands we consider BSD analytics as just one specific modality of data motility—that of contained mobility. For indeed, this data flows through corporate enclosures, in a manner *not* directed by the digital human who generated it. But in critically unpacking this contained data mobility we need to consider the breadth of the heterogeneous ensemble through which it flows, to discover other intentionalities and desires which may indicate more liberatory possibilities of data motility.

Acxiom is a little known but major American data broker that collects consumer data, information from financial service companies, court records and government documents. As recently outlined by the Electronic Freedom Foundation, they have partnered with Facebook.[36] For example, Facebook will identify a desired audience, say potential car buyers. Acxiom will then scour its databases and create a list of everyone who meets that criteria and provide it to Facebook. That list will then be delimited by Facebook to include only its users which in turn will be served up to the car manufacturer so it can effectively produce appropriate ads. Finally, Facebook

will display that ad alongside the targeted user's newsfeed. There are a number of things worth emphasising in this example of data motility. For one, it highlights the ever-multiplying stages of motility, of the movement of the data we create but do not direct. First, the digital human generates the structured data of government records, financial documents and consumer behaviour. Second, this data moves from its initial database to those of Acxiom. Third, these discrete elements are moved again at the behest of Facebook, in aggregation by Acxiom. Fourth, they are collectively moved again to Facebook. Fifth, they move from Facebook to the auto manufacturer. Sixth, the discrete points of data users once generated, now profoundly processed and aggregated, are pinged back to those same digital humans in the form of a targeted ad. Finally, user response to those targeted ads become a new source of BSD in a deep layer of recursivity: 'Facebook then provides the company with an aggregate report about how an ad performed, which might include information about how many people clicked on it, their locations, ages, genders, etc.'[37]

This latter point leads to the next generation of BSD analytics that Facebook is unveiling, in a formal partnership with Acxiom, Datalogix, Epsilon and Bluekai. Acxiom, like Datalogix and Epsilon, has its own databases, culled from loyalty cards, purchase-based data and other comprehensive demographic databases. BlueKai, however, contributes uniquely to an even more heterogeneous and frictionless flow of BSD, specialising in tracking cookies which collect information about all the sites you visit when not on Facebook. Upon returning, an HTML pixel web bug enables Facebook to process the data about all the other sites you visited. This provides the social media giant with a comprehensive digital trace of your online predilections, which, in turn, can be analysed and aggregated with all the aforementioned data now in their proprietary grasp. This 'cookie matching' makes you even more valuable for advertisers who want to target you on Facebook.

To facilitate this next stage of heterogeneous BSD integration, Facebook has purchased Atlas, an ad-server formerly owned by Microsoft. As *Advertising Age* notes, this is a clear sign of Facebook's intention to become an online ad server behemoth, second only to Google's DoubleClick.[38] First, this new ad server will consolidate advertiser connection to Facebook's display tools and exchange, and to the subsequent measurement of onsite advertisement effectiveness. But given the increasingly integrated and heterogeneous flow of information and collection points,

the quantification of effectiveness is no longer limited to whether or not you click on the targeted ad. Your consumer habits can now be tracked outside your Facebook-based activities via the myriad databases of the new array of partners; for example, via your general online habits or when you use your credit card. In turn, this can be analysed via forms of textual analysis of the user content you generate on Facebook. These kinds of sophisticated BSD analytics facilitate a particular kind of data motility which seeks to quantify the affective sociality of advertising. That is, it hopes to measure not just your click through rate but the impression of ads. As Mark Zuckerberg stated to investors, the strategic intention is to 'help connect ad impressions and purchases'.[39]

These specific material developments and configurations facilitate the ever-more comprehensive capture of data for corporate purposes. There are also regulatory decisions and laws enabling a more frictionless flow between those companies and the state. For example, the US House of Representatives recently passed by a wide margin the Cyber Intelligence Sharing and Protection Act (CISPA). If enacted, this bill would allow companies to monitor user actions that leave a digital trace and share it with the government, without a warrant and without ever needing to notify you that it possesses your data, regardless of how sensitive it might be. 'This means a company like Facebook, Twitter, Google, or any other technology or telecoms company, including your cell service provider, would be legally able to hand over vast amounts of data to the US government and its law enforcement—for whatever purpose it deems necessary—and face no legal reprisals.'[40] Further, such state compulsion to share data without consent or knowledge would not be subject to the Freedom of Information Act which otherwise would enable the public to request that government release information.

It must be stressed that at the time of writing, this bill remains in legislative limbo, with the US senate refusing to vote on it because of concerns over insufficient privacy protection, and because of political infighting resulting from National Security Agency (NSA) revelations. Nonetheless, there are other examples around the world. India has invoked the Central Monitoring System, which will allow the government and its agencies to monitor all telecommunications and internet communications within that country. According to the Centre for Internet and Society this enables a general environment of e-surveillance that establishes central

and regional databases and allows central and state law enforcement agencies to intercept and monitor communication, as well as to undertake call data record analysis and data mining.[41]

The rise of such new regulations across the globe, and the disturbing practices of NSA data capture and analysis, indicate the need for critical debate around privacy in the age of BSD. Such new laws are justified by the purported need for cybersecurity. These are key issues in need of informed consideration but are beyond the scope of this article. Instead, I want to posit this less in discursive or ideological terms of security and more as an effect of the material elements of the *dispositif* of data motility. Given the persistence and permanence of our broadly generated digital traces, and the material changes enabling the intensive and extensive processing of different forms of data, there should be no surprise that an ever-more frictionless flow becomes an urgent need for both the state and capital. These common interests are clearly visible on the surface. Lobbyists in favour of CISPA outspent opponents by 140 times, and include major tech, telecommunication and financial corporations, including AT&T, Comcast, Verizon, Time Warner Cable, National Cable and Telecommunications Association, Cellular Telecom & Internet Association, Oracle, Intel, IBM and the American Bankers Association.[42] To suggest this group is a cabal that planned and orchestrated the widespread and frictionless flow of BSD is to miss the point of a *dispositif*. Rather, look to the cohesion, the binding of strategic interests under the logic of data that can retain its information as it moves and is processed in myriad and ongoing iterations.

There is one more potential regulatory change that must be mentioned. The precise articulation of property rights calibrates the control exercised over the flow of data. Intellectual property law and user agreements are key regulations that guarantee the controlled flow of BSD through a highly proprietary environment. In a social media context, one owns the data one generates, insofar as a copy can be demanded from Facebook. That does nothing, however, to limit the secondary rights held by the social media giant, which moves, mines, processes and aggregates your data at will. The status of data ownership in a cloud environment was brought further into question with the FBI-led case against Megaupload. When Megaupload's servers, holding about twenty-five petabytes of data, were unplugged last year, the data property rights of those using Kim Dotcom's services were seemingly

abrogated. One such user, Kyle Goodwin, used Megaupload to store video and files for his small regional website that covers high school sports. He has to date unsuccessfully sought to retrieve his data, subsequently taking legal action based on the argument that the US government, in its pursuit of Megaupload, had not taken reasonable steps to protect third-party property rights in cloud computing storage.

The US government has strongly opposed Goodwin's efforts. According to Goodwin's lawyers: 'Apparently your property rights "become severely limited" if you allow someone else to host your data under standard cloud computing arrangements.'[43] Further, even if the government's position does not stand up to legal challenge, they have indicated they will implement administrative measures whereby the data would first need to be reviewed by the government or a third party to determine if any of it infringed copyright. It is worth noting that the Motion Picture Association of America has filed a brief as a non-party participant in the case, in support of that system.[44] These examples from Facebook and Megaupload demonstrate the prominence of data mobility as a modality of control, surveillance and profit, and cannot be underestimated. But what remains in the *dispositif* of data motility?

—FOR A DATA DEBT JUBILEE?

For the *dispositif* to be a sharp tool for critical analysis, its heterogeneity must be foregrounded, both in terms of its discursive and material elements and in the differentiated power and knowledge relations it engenders. The Hadoop material structure does not require a proprietary environment. It is the strategic interests of big data and social media companies that results in the parsing of data for a controlled flow. Yet there is nothing in the material environment of BSD that leaves it exclusively bound to an algorithmic power of profitable and productive control. Just as biopower's *dispositif* of control and domination must be differentiated from the biopolitical domain of creativity and resistance, a similar distinction must be made between data mobility and data motility. I suggest differentiating the proprietary environment as one of *mobility*, wherein the flow of data is motile vis-à-vis its being wholly autonomous of the control of those who generated it, but is ultimately directed by social media and big data companies which calibrate its flow

for maximum profitability. Indeed, it could be stated that the state and capital embrace controllable data mobility but fear and loathe autonomous data motility.

Let's go back to the material phenomenon that inspired the conceptualisation of BSD through the *dispositif* of data motility. One of the defining features of cloud architecture was the virtual disappearance of physical boundaries containing your data. There are, of course, very clear material boundaries that remain, but they can be literally distributed across the globe. As well, the cloud environment is typically a shared one, and the vicissitudes of data optimisation require a replication factor of at least three, meaning that each unique 'raw data event' is stored in at least three locations across the cloud. Further, this is dynamic data replication, so your 'raw data event' could be in the northern hemisphere one moment and in the southern the next. Finally, the movement of this data between geographically distributed data centres regularly happens with neither administrator knowledge nor consent. This is a structural glitch in the cloud wherein data moves autonomously, in a seeming act of self-generated movement. Motility is, above all else, autonomous movement. In this specific instance, data motility is a material expression of the cloud's architecture and code. One data security expert bemoaned the strange phenomenon whereby cloud-stored data moves of its own accord, complaining of 'the headaches that come from unruly and nomadic information'.[45] The literal source of data motility, then, is strictly a material effect of cloud architecture. What is of greater critical analytical value and potential political import, however, are the implications of a cloud that 'leaks' data. Indeed, the same data security expert expressed concerns in surely unintended Deleuzo-Guattarian terms: 'unruly and nomadic'. In the remainder of this article, I introduce for further consideration some potentially deterritorialising effects that may result from the motility of data.

BSD, then, is literally motile in cloud storage, and it is proprietarily so with the big data and social media companies for whom it is a motile commodity. I have illustrated how the data we create moves autonomously of our control and constantly 'pings' back to us in ways that delineate the topology of our everyday life. Yet there is a deeper felicity in motility—which invites further study—because it links the autonomous material movement of BSD to a rich and varied philosophical tradition wherein *motility* is the ontological baseline for *Being*. Aristotle, for example, places *kinesis* (later translated as motility) at the centre of philosophical

reflection. Significantly, Aristotle limited motility to natural things (*phusis*), and that has long demarcated the natural from the artificial, the human from the nonhuman. Elsewhere I have written at length about the constitutive relationship between the human and technology, which I present here as the digital human.[46] I ground my understanding of the human-nonhuman assemblage in Leroi-Gourhan's concept of originary technicity, which has subsequently influenced both Derrida and Stiegler, among others.[47] The core claim made by Leroi-Gourhan is that human speciation occurred in a deeply recursive relationship with technology. This startling claim, that we have only ever been human in an assemblage with technology, undermines much of Western metaphysics, beginning with Plato and Aristotle, and calls into question that strict separation of the natural human and artificial technology. The digital human is a means for thinking of the human as always already being in a constitutive assemblage with technology. The specific material elements of that assemblage are of great importance, as they indicate the importance of historicising those mediating elements.

To suggest that the assemblage's material, technologically mediated elements—the nonhuman, as it were—have gained proper motility highlights what is unique about the digital human. Motility was also key to Hegelian logic wherein dialectics turn on the Being of life in its specific motility. Heidegger thought motility constituted as opposed to something that happens to being. Here we should pause to think of the implications of originary technicity, wherein humanisation begins with the exteriorisation of memory into rudimentary stone tools. BSD, in this sense, is nothing but the exteriorisation of memory, of the quotidian, mediated actions of life. The motility of our BSD is not something that happens to us; it is constitutive of our being as digital humans. Keith Ansell Pearson provocatively reads Heideggerian motility in terms sympathetic to this perspective, positing a Deleuzian ethology, wherein it signifies the becoming of life but only ever in a deeply relational structure with 'environment'.[48] Finally, Marcuse posits the motility of being as the historicising rootedness in the world, linking it to both labour, and radical acts of social, political and economic transformation. I put forth motility, then, because it denotes a potentially transformative becoming in a deeply recursive and historicised mediated environment. As such, data motility marks the tensions and struggles endemic to the digital human, and is in need of further critical inquiry. Just

how the quotidian data that we generate moves autonomously of our control circumscribes the ontological ground of the digital human. In the space that remains, I want to suggest possible ontological implications made visible by the *dispositif* of data motility, specifically as it relates to life and labour under BSD.

If we return to the *dispositif*, then, the assemblage of data motility resonates deeply with Ansell Pearson's reading of Heideggerian motility. The environment with which the becoming of life is relational, is conducive to the generation and intensive processing of BSD. One interpretation is that the proprietary environment of control engendered by data motility is also one wherein *Being* is in a state of data encumbrance. Here we can turn to Lazzarato, noted already for explicating the polyvalent nature of the *dispositif*, through the example of biopower-biopolitics, which in turn I am applying to mobility-motility. In his recent book, *The Making of Indebted Man*, Lazzarato extends his critique of 'immaterial labour' which denotes the increasingly prominent role of communicative sociality in the generation of capitalist value. His thesis is that the debtor-creditor relationship is the core of the neoliberal condition.

I find this suggestive, particularly in terms of data motility. Lazzarato contends that debt breaks down the binaries producer–consumer and working–nonworking. He sees this as a radical extension of biopower wherein debt is a strategy of control, a rearticulation of its imperative 'become productive'. He posits 'indebted man' (*sic*) as the subjective figure of contemporary capitalism: 'Debt breeds, subdues, manufactures, adapts, and shapes subjectivity. What kind of subjectivity?'[49] Lazzarato situates debt as a correlate of Deleuzian control society, as opposed to the confinement of disciplinary society. This provides an interesting counterpoint to an important body of work that situates social networks and Web 2.0 environments in terms of data enclosure and confinement.[50] But how might debt be applied to the material environment of BSD? One way is to see a command of encumbrance, a strategy of control exercised through the *dispositif* of data motility. Let us recall how the conflation of consumer–producer and work–nonwork are hallmarks of the social web, of social media, and of immaterial labour 2.0. The rise of BSD takes us further along this continuum, as a more extensive and intensive variant.

Let me try to nuance this claim, to outline an approach to further study. I suggest that BSD comprises the endless payments we make to neoliberal digital or

cognitive capitalism. In order to access any social media platform, any element of Web 2.0, we must generate social data. It is structurally unavoidable, and the motility of that data is the means by which its sociality is turned into economic value. This renders BSD as a key modality for responding productively to the command of neoliberal debt. As Lazzarato emphasises throughout his recent work, debt encourages and compels us to become the 'entrepreneurs' of ourselves, as 'human capital'.[51] The capital of the digital human is data. Data—as metadata and user-generated content—is highly productive for capital, given its strategy to buy it low and sell high. This dynamic of debt runs through the *dispositif* of data motility. Social media, be it Facebook, Twitter or Google, is on the surface free for users. In turn, content is generated for free. The entire business model of social media platforms turn on selling that data as profitably as possible. Hence the growing appeal of the Hadoop environment, of the intensive and ongoing processing of BSD. It is an environment structured to maximise data motility wherein data moves autonomously of your control from the moment you generate it.

Yet, as already noted, data motility is not just a *dispositif* of control that harnesses the digital traces of life for work; it is equally one that offers new political and economic opportunities for constituent power and resistance. Motile data is social data, and the *sociality* of that data highlights its polyvalence—the social *and* economic valorisation that underpin social media. Sociality is the driver of BSD. These are new mediated cultural practices, and the resulting BSD is generated by social, communicative and affective relations. They are transformed into economic relations, as noted by Wostall, the fellow from the Adam Smith Institute (which, it should be recalled, was the intellectual force behind privatisation under UK prime minister Margaret Thatcher). The circulation, exchange and valuation of such interlinked social data is crucial to the expansion of neoliberal digital capital. Nonetheless, it is the sociality of data, not the strategies of its capture, that coheres the *dispositif* of data motility. Attention to the materiality of the *dispositif* of data motility, further, indicates that he is right that 'we' are being 'sold' in social media. I find it far more interesting, however, to regard this not just as yet another normative capitalist relation, but also as a new form of debt that encumbers the breadth and depth of our newly gained communicative and social capacity. When viewed this

way, data motility signals concomitant possibilities of new digital commons and political action.

In this regard it seems nonsensical, as political strategy, to try and strip ourselves of BSD. There is a profound potential therein for expanded and intensified communicative and affective capacities. As Pybus notes, the archive of BSD, as a kind of archive of everyday life, is not merely the sedimented part but also a liminal space.[52] What seems intolerable is the prospect of it remaining a space for becoming a more profitable consumer or a better surveilled subject. What a critical understanding of the *dispositif* of data motility helps clarify is that collective sociality comes before its capture by capital. Here we benefit from recalling that sociality drives the desiring assemblage of motility. This helps us think about ways to reject our data encumbrance and to reclaim what, after all, is ours. What might a radical embrace of data motility mean? What are the algorithmic codes that can create libidinal economies from a new data commons? What might a BSD commons look like? What kind of new sociality might emerge in critical projects of personal data curation? What are the political possibilities that data motility—which seems inherently deterritorialising—hold for, among other things, 'the exploit' about which Galloway and Thacker write so provocatively?

Lets give the last word to an emerging player in the Hadoop environment, Platfora, which was recently bolstered by major investment, including from InQTel, the CIA's venture capital arm. Platfora seeks to make BSD open to real-time intuitive, and serendipitous analysis for its corporate clients. 'Imagine what is possible … [when e]veryday business users can interactively explore, visualise and analyse any of that data immediately, with no waiting for an IT project. One question can lead to the next and take them anywhere through the data.'[53] What might be possible if such analyses were taken up by community organisations, affinity groups, hackers and radical political movements? Such new power and knowledge relations are possible under the dispositif of data motility.

'Imagine what is possible.'

—

Mark Coté is now program director of the Masters in Digital Culture and Society at King's College London and a member of both the Department of Digital Humanities and the Department of Culture, Media and Creative Industries. He has a sustained research trajectory in social media, digital culture and the relationship between the human and technology. He has written extensively on digital culture via Foucauldian and Italian post-operaismo/autonomist thought with an increasing focus on the materiality of the digital ecosystem in which we live. He is currently co-investigator on an AHRC-funded research on the 'big social data' we generate on our mobile devices.

—NOTES

[1] Richard Barbrook and Andy Cameron, 'The Californian Ideology', *Science as Culture*, vol. 26, 1996, pp. 44–72.

[2] Tim Worstall, 'Facebook Is Free Therefore It Is You Getting Sold', *Forbes*, 11 October 2012, <http://www.forbes.com/sites/timworstall/2012/11/10/facebook-is-free-therefore-it-is-you-getting-sold/>.

[3] Mark Coté and Jennifer Pybus, 'Learning to Immaterial Labour 2.0: MySpace and Social Networks', *Ephemera*, vol. 1, no. 7, 2007, pp. 88–106.

[4] Michel Foucault, 'From Torture to Cellblock' in *Foucault Live: Collected Interviews*, *1961–1984*, ed. S. Lotringer, trans. J. Johnston, Semiotext(e), New York, 1996 pp. 146–9.

[5] Gilles Deleuze, 'Desire and Pleasure' in *Foucault and his Interlocutors*, ed. A. Davidson, trans. Daniel W. Smith, University of Chicago Press, Chicago, 1994, pp. 183–92; Gilles Deleuze, 'What is a Dispositif?' in *Michel Foucault, Philosopher: Essays*, ed. Timothy J. Armstrong, trans. Timothy J. Armstrong, Harvester Wheatsheaf, London, 1992, pp. 159–68.

[6] Michel Callon, 'Why Virtualism Paves the Way to Political Impotence: A Reply to Daniel Miller's Critique of The Laws of the Markets', *Economic Sociology: European Electronic Newsletter* 6/2, February 2005, pp. 3–20.

[7] Rick Dolphijin and Iris van der Tuin, *New Materialism: Interviews & Cartographies*, Open Humanities Press, Ann Arbor, 2012.

[8] Deleuze, 'Desire and Pleasure'.

[9] There is a wide range of work over the past several years that has focused on the infrastructural or material elements of social media, Web 2.0. A very brief list would include the following: Jean-François Blanchette, 'A Material History of Bits', *Journal of the American Society for Information Science and Technology*, vol. 62, no. 6, 2011, pp. 1042–57; Tainia Bucher, 'Want to be On the Top? Algorithmic Power and the Threat of Invisibility on Facebook', *New Media & Society*, vol. 14, no. 7, 2012, pp. 1164–

80; Kevin Driscoll, 'From Punched Cards to "Big Data": A Social History of Database Populism', *communication +1*, vol. 1, no. 4, 2012, <http://scholarworks.umass.edu/cpo/vol1/iss1/4>; Ganaelle Langlois, and Greg Elmer, 'Wikipedia Leeches? The Promotion of Traffic through a Collaborative Web Format', *New Media & Society*, vol. 11, no. 5, 2009, pp. 773–94; Sabine Niederer and Jose Van Dijck, 'Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System', *New Media & Society*, no. 12, vol. 8, 2010, pp. 1368–87.

[10] Michel Foucault, 'Le jeu de Michel Foucault' in *Dits et écrits: Tome 3*, eds D. Defert and F. Ewald, Gallimard, Paris, 2001, pp. 298–329.

[11] Ibid.

[12] Michel Foucault, 'Les Mailles du pouvoir' in *Dits et écrits Tome IV*, eds D. Defert and F. Ewald, Éditions Gallimard, Paris, 1994, pp. 182–201.

[13] Maurizio Lazzarato, 'From Biopower to Biopolitics', *Pli: The Warwick Journal of Philosophy*, vol. 13, 2002, pp. 112–25.

[14] Deleuze, 'What is a Dispositif?'

[15] See Jennifer Pybus, 'Social Networks and Cultural Workers: Towards an Archive for the Prosumer', *Journal of Cultural Economy*, vol. 6, no. 2, forthcoming 2014.

[16] Michel Foucault, *Archaeology and Knowledge*, trans. A.M. Sheridan, Tavistock, London, 1972.

[17] Richard L. Villars, Carl W. Olofson and Matthew Eastwood, 'Big Data: What It Is and Why You Should Care', *International Data Corporation White Paper*, 2011, <http://sites.amd.com/us/Documents/IDC_AMD_Big_Data_Whitepaper.pdf>.

[18] John Gantz, and David Reinsel, 'The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East', *IDC iView*, 2012, <http://www.emc.com/leadership/digital-universe/iview/index.htm>.

[19] Wikibon Blog, 'A Comprehensive List of Big Data Statistics, 2012', <http://wikibon.org/blog/big-data-statistics/>.

[20] A 1998 Merrill Lynch study estimated more than 80 per cent of all data was unstructured, a figure that since has been widely cited. Information management giants, from IBM to IDC to Gartner to Forrester, all use this figure, and estimate that over the next five years 80 per cent of new data created will be unstructured. See Adrian Bridgewater, 'IBM: 80 Percent of Our Global Data is Unstructured (so what do we do?)', *ComputerWeekly*, 26 October 2010, <http://www.computerweekly.com/blogs/cwdn/2010/10/ibm-80-percent-of-data-is-unstructured-so-what-do-we-do.html>; and Chris Preimesberger, 'Managing Massive Unstructured Data Troves: 10 Best Practices', *eWeek*, posted 3 July 2013, <http://www.eweek.com/storage/slideshows/managing-massive-unstructured-data-troves-10-best-practices/>.

[21] Alexander Galloway, *Protocol*, MIT Press, Cambridge MA, 2004; Robert Gehl, 'Real (Software) Abstraction', *Social Text*, vol. 30, no. 2, 2012, pp. 99–119; and Peter Mika, *Social Networks and the Semantic Web*, Springer, New York, 2007.

[22] Foucault, 'Le jeu de Michel Foucault'.

[23] Ibid.

[24] Jeffery Dean and Sanjay Ghemawat, 'MapReduce: Simplified Data Processing on Large Clusters', *Proceedings of the 6th Symposium on Operating Systems Design and Implementation*, San Francisco, 6–8 December 2004, <http://static.usenix.org/event/osdi04/tech/full_papers/dean/dean_html/>.

[25] S. Ghemawat, H. Gobioff and S.T. Leung, 'The Google File System', *ACM SIGOPS Operating Systems Review*, vol. 37, no. 5, 2003.

[26] Dean and Ghemawat.

[27] Joseph Conrad, *Lord Jim*, Broadview Press, Peterborough, 2000. pp. 283.

[28] Amr Awadallah, 'Introducing Apache Hadoop: The Modern Data Operating System' (video), Stanford Center for Professional Development, Stanford, 2011, <http://www.youtube.com/watch?v=d2xeNpfzsYI>.

[29] Matthew Komorowski, 'A History of Storage Cost', <http://www.mkomo.com/cost-per-gigabyte>.

[30] Amr Awadallah, 'Hadoop: Distributed Data Processing', 27 January 2010, <http://www.sfbayacm.org/wp/wp-content/uploads/2010/01/amr-hadoop-acm-dm-sig-jan2010.pdf>.

[31] Consider this computer industry-based perspective from a mere decade ago bemoaning the persistent difficulties in processing different types of data: Robert Blumberg and Shaku Atre, 'The Problem with Unstructured Data', *Information Management*, 1 February 2003, <http://soquelgroup.com/Articles/dmreview_0203_problem.pdf>.

[32] Hadoop Wiki: PoweredBy, <http://wiki.apache.org/hadoop/PoweredBy>.

[33] Andrew Ryan, 'Under the Hood: Hadoop Distributed Filesystem Reliability with Namenode and Avatarnode', *Facebook Engineering (notes)* 2012, <https://www.facebook.com/notes/facebook-engineering/under-the-hood-hadoop-distributed-filesystem-reliability-with-namenode-and-avata/10150888759153920>.

[34] Doug Henschen, 'Facebook on Big Data Analytics: An Insider's View', *InformationWeek Cloud*, 18 March 2013, <http://www.informationweek.com/cloud-computing/platform/facebook-on-big-data-analytics-an-inside/240150902>.

[35] Mark Andrejevic, 'Social Network Exploitation' in *Networked Self: Identity, Community and Culture on Social Network Sites*', ed. Zizi Papacharissi, Routledge, New York, 2011, pp. 82–103. Christian Fuchs, 'New Media, Web 2.0 and Surveillance', *Sociology Compass*, vol. 5 no. 2, 2011, pp. 134–47.

[36] Kurt Opshal and Rainey Reitman, 'The Disconcerting Details: How Facebook Teams Up With Data Brokers to Show You Targeted Ads', *Deeplinks Blog, Electronic Frontier Foundation*. 22 April 2013,

<https://www.eff.org/deeplinks/2013/04/disconcerting-details-how-facebook-teams-data-brokers-show-you-targeted-ads>.

[37] Opshal and Reitman.

[38] Jason Del Ray, 'Facebook Lays Groundwork For Ad Network Off The Social Site', *Advertising Age*, 11 February 2013, <http://adage.com/article/digital/facebook-positioned-facebook-ad-net-atlas-buy/239713/>.

[39] Tomio Geron, 'Live: Mark Zuckerberg On Facebook Q1 Earnings Call', *Forbes*, 1 May 2013, <http://www.forbes.com/sites/tomiogeron/2013/05/01/live-mark-zuckerberg-on-facebook-q1-earnings-call/>.

[40] Zack Whittaker, 'CISPA "Dead" in Senate, Privacy Concerns Cited', *ZDNet*, 25 April 2013, <http://www.zdnet.com/cispa-dead-in-senate-privacy-concerns-cited-7000014536/>.

[41] Maria Xynou, 'India's Big Brother: The Central Monitoring System (CMS)', *The Centre for Internet & Society*, 8 April 2013, <http://cis-india.org/internet-governance/blog/indias-big-brother-the-central-monitoring-system>.

[42] Keenan Steiner and Bob Lannon, 'Pro-CISPA Forces Spend 140 Times more Lobbying than Opponents', *Sunlight Foundation Reporting Group*, 18 April 2013, <https://data.sunlightlabs.com/dataset/Lobbying-totals-by-CISPA-proponents/5brg-ruk9>.

[43] Cindy Cohn and Julie Samuels, 'Megaupload and the Government's Attack on Cloud Computing', *Deeplinks Blog*, 31 October 2012, <https://www.eff.org/deeplinks/2012/10/governments-attack-cloud-computing>.

[44] Declan McCullagh, 'MPAA: No MegaUpload Data Access without Safeguards', *C/Net*, 30 October 2012, <http://news.cnet.com/8301-13578_3-57542707-38/mpaa-no-megaupload-data-access-without-safeguards/>.

[45] Dan Crowe. 'Did You Know … In the Cloud, Your Data Can Get Up and Move on its Own', *Cloud Security Blog*, 28 October 2010, <http://cloud.trendmicro.com/did-you-know%E2%80%A6-in-the-cloud-your-data-can-get-up-and-move-on-its-own/>.

[46] Mark Coté, 'Technics and the Human Sensorium: Rethinking Media Theory Through the Body', *Theory & Event*, vol. 13, no. 4, 2010; Mark Coté, 'The Pre-Historic Turn?: Networked New Media, Mobility and the Body' in *The International Companions to Media Studies: Media Studies Futures*, ed. Kelly Gates, Blackwell, Oxford, 2012, pp. 171–94.

[47] Jacques Derrida, *Of Grammatology*, Johns Hopkins University Press, Baltimore, 1976; Bernard Stiegler, *Technics and Time, 1: The Fault of Epimetheus*, Stanford University Press, Paolo Alto, 1998.

[48] Keith Ansell Pearson, *Viroid Life: Perspectives on Nietzsche and the Transhuman Condition*, Routledge, New York, 1997.

[49] Maurizio Lazzarato, *The Making of Indebted Man*, Semiotext(e), New York, 2012, pp. 38–9.

[50] For two prominent examples, see Mark Andrejevic, 'Privacy, Exploitation, and the Digital Enclosure', *Amsterdam Law Forum*, vol. 1, no. 4, 2009, and Greg Elmer, 'A Diagram of Panoptic Surveillance', *New Media & Society*, vol. 5, no. 2, 2003, pp. 231–47.

[51] Lazzarato, *The Making of Indebted Man*, pp. 50.

[52] Jennifer Pybus, 'Social Networks and Their Archives of Feeling' in *Networked Affect*, eds S. Paasonen, K. Hillis and M. Petit, MIT Press, Cambridge, MA, forthcoming 2014.

[53] Ben Werther, 'The End of the Data Warehouse', *Platfora Blog*, 23 October 2012, <http://www.platfora.com/the-end-of-the-data-warehouse/>.