

# Data Out of Place

## Toxic Traces and the Politics of Recycling

Thylstrup, Nanna Bonde

*Document Version*  
Final published version

*Published in:*  
Big Data & Society

*DOI:*  
[10.1177/2053951719875479](https://doi.org/10.1177/2053951719875479)

*Publication date:*  
2019

*License*  
CC BY-NC

*Citation for published version (APA):*  
Thylstrup, N. B. (2019). Data Out of Place: Toxic Traces and the Politics of Recycling. *Big Data & Society*, 6(2).  
<https://doi.org/10.1177/2053951719875479>

[Link to publication in CBS Research Portal](#)

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us ([research.lib@cbs.dk](mailto:research.lib@cbs.dk)) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Sep. 2024



# Data out of place: Toxic traces and the politics of recycling

Big Data & Society  
July–December 2019: 1–9  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2053951719875479  
journals.sagepub.com/home/bds



Nanna Bonde Thylstrup 

## Abstract

It has become increasingly common to talk about “digital traces”. The idea that we leak, drop and leave traces wherever we go has given rise to a culture of traceability, and this culture of traceability, I argue, is intimately entangled with a socio-economics of data disposability and recycling. While the culture of traceability has often been theorised in terms of, and in relation to, privacy, I offer another approach, framing digital traces instead as a question of waste. This perspective, I argue, allows us to connect to, extend and nuance existing discussions of digital traces. It shows us that data *traces* raise questions about not only how data capitalism *tracks* individual and multiple data behaviours, but also how it links to social and environmental *toxicities* in the form of abuse and environmental pollution, which follow gendered and colonial structures of violence.

## Keywords

Digital traces, knowledge production, Big Data ecologies, toxic data, recycling, waste

This article is a part of special theme on Knowledge Production. To see a full list of all articles in this special theme, please click here: <http://journals.sagepub.com/page/bds/collections/knowledge-production>

## Big Data: A matter of waste

Datafication is often framed as an innovative form of knowledge production that distils ‘old’ information into pure data points that can then be repurposed for new insights (Flyverbom and Murray, 2018; Kitchin, 2014; Ruppert et al., 2017). While the emphasis is often put on novelty in matters of digitisation, this article argues that the logic of datafication is premised on a logic of waste and recycling, with significant implications for how we consider datafication’s politics and ethics. To support my argument I draw on the insights of discard studies, first to unpack the concept of digital traces as a question of waste, second to link digital traces to an economic logic of extraction, and third to show how the data culled from these processes of extraction is never ‘pure’ but rather marked by the traces of the bodies from whence it came.

We arguably live in a new culture of traceability. Platforms accumulate and distribute subjects’ traces, using machine learning processes to construct

histories – and ultimately patterns – of action (Chun, 2016b). The concept of digital traces thus signifies more than pure data points (Hepp et al., 2018: 2); rather, digital traces raise questions about how we make meaning with Big Data. Understanding datafication as a question of meaning-making attunes us to two major claims: that the knowledge produced by Big Data is also marked by relations of power; and that these relations constitute a *system* of representation, which consists of different ways of organising, clustering, arranging and classifying concepts, and establishing complex relations between them (Hall, 2013). Within this system, we find a constant recycling of digital traces that extracts and repurposes previous

Det Humanistiske Fakultet, Københavns Universitet, København, Denmark

### Corresponding author:

Nanna Bonde Thylstrup, Det Humanistiske Fakultet, Københavns Universitet, København 2300, Denmark.  
Email: [nbt.msc@cbs.dk](mailto:nbt.msc@cbs.dk)

forms of meaning, embedding them in new representations and ‘dramas’ of Big Data (Chun, 2016a). Big Data knowledge production, as Lisa Blackman (2020) notes, is therefore a question of *invention* as much as discovery.

The discourse of digital traces plays a special role in today’s digital economy, which largely bases itself on the assumption that the knowledge gleaned from the digital traces left behind by users (intentionally or unintentionally) translates into power in the marketplace. Digital traces have become ‘a primary resource for value creation, influence, and knowledge production’ in Big Data (Flyverbom and Murray, 2018: 2), because they can help ‘internet companies know users, gain insights about customer preferences and design new products and markets’ (Flyverbom and Murray, 2018: 6). Investments are therefore made to collect and extract even more digital traces from users of platform-based online services. At the same time, however, the close monitoring, collection and analysis of digital traces has also given cause for concern, resulting in new user behaviours that seek to minimise users’ digital traces in online spaces – for instance, through ‘detoxing’ and other practices of disconnectivity (Hesselberth, 2018), as well as new legal measures such as the General Data Protection Regulation to counter predatory and exploitative behaviour by digital companies. Yet even these evasion strategies, which seek to provide some form of privacy, produce digital traces, as Mayer-Schönberger and Cukier (2014: 154) note: ‘if everyone’s information is in a dataset, even choosing to “opt out” may leave a trace’. (Hesselberth (2018) makes the same point, arguing that there is no disconnectivity without connectivity.)

As Sarah Myers West (2017: 2) points out, the commodification of digital traces introduces a logic of data capitalism that ‘places primacy on the power of networks by creating value out of the digital traces produced within them’. The logic of data capitalism thus shares clear affinities with Shoshana Zuboff’s (2019) concept of surveillance capitalism, which posits a new subspecies of capitalism that profits from the combined surveillance and modification of human behaviour.

While the issue of traceability and its implications for how we theorise knowledge production have been treated at length in the realm of privacy theory and dataveillance (Agre, 1994; Cohen, 2012; Mai, 2016; Nissenbaum, 2010), producing insightful work on the ways in which our actions are tracked, captured, transformed into (big) data and made actionable by companies and states (e.g. Lyon, 2008; Zuboff, 2019), this article offers a different perspective on digital traces: through cultural theories of waste and recycling.

The contribution of discard studies, as it is formulated by feminist and anti-colonial scholars such as Josh Lepawsky (2018) and Max Liboiron (2016; Liboiron et al., 2018; Jennifer Gabrys (2010) and Jussi Parikka (2011) among many others offers not only a theoretical grounding for the already pervasive metaphorical presence of waste in critical data studies (e.g. ‘garbage-in-garbage-out’, data exhaust, dirty data, bit rot, toxic data, data sweat and data trash), but also a way of thinking about approaches to broader systems of power in Big Data knowledge production and their complex ecological systems.

Cultural theories on waste and discard, I suggest, allow us to connect to, extend and nuance existing discussions of digital traces. They show us that data *traces* raise questions not only about how data capitalism *tracks* individual and multiple data behaviours, but also about how it links to social and environmental *toxicities* in the form of abuse and environmental pollution, which follow gendered and colonial structures of violence. Rather than eliminating the displaced and suppressed narratives of these structures, waste theories show us that datafication is instead always *haunted* by these traces and the threat that they will resurface.

## The concept of digital traces

The notion of digital traces has become a common way to describe what we leave behind online. As such, it constitutes both a cultural metaphor and a material externality. According to the *Oxford English Dictionary*, ‘trace’ can be taken to designate ‘a non-material indication or evidence of the presence or existence of something, or of a former event or condition; a sign, mark’ and also a very material ‘track made by the passage of any person or thing, whether beaten by feet or indicated in any other way’ (OED, 2019). The purpose of this section is to explore both the cultural work and the material externalities of this metaphor. In its attention to the materiality of the digital trace, this article joins current efforts in critical media studies to explore the environmental costs of digitisation and datafication (Cubitt, 2014; Gabrys, 2010, 2016; Hogan, 2015; Parikka, 2011).

If the concept of data appears to describe a wholly immaterial phenomenon that does not engage the senses, the concept of digital trace adds a more particular signification as something that is intentionally or unintentionally left behind. In digitisation, this signification has come to signal a way of viewing and treating data as a form of material that can be gleaned, mined and put to new uses, irrespective of the system to which that data previously belonged. Data appears as a processed good that may nevertheless be mined as raw

material to make it give up its meaning (Gitelman, 2013; Räsänen and Nyce, 2013).

As one book on computational methods states, ‘unlike footprints in the sand, *digital traces* in silica are not wiped away by the tide; instead they accrete, leaving behind *incredibly detailed records of social interaction*’ (Welser et al., 2010: 117). Digital traces have become a coveted asset in many fields of cultural analysis, because they are framed as providing ‘unobtrusive measures of people’s thoughts at a given point of time’ (Alexander et al., 2018: 2). This perceived quality also makes them prized commodities. As Sarah Myers West (2017: 2) notes, the traces generated by our daily lives are ‘collected, aggregated, fed into algorithms, and used to predict our behavior for a variety of purposes: to sell advertisements, certainly, but also to calibrate technologies, improve search results, contribute to valuable research, and more nefariously, to feed intelligence agencies’ insatiable appetite for knowledge about our global communications’. While datafication has given rise to a new ‘science of traceability’ (Bigo, 2006: 60), researchers have ‘always relied on media inscriptions to investigate collective phenomena’ (Venturini et al., 2018). In that sense, digital traces belong to already existing dynamics of knowledge production (Boullier, 2017). Yet datafication has also accelerated and distributed the means of traceability exponentially, thereby amplifying and modulating processes of inscription and tracking (Venturini et al., 2017: 2).

Contemporary discourses on these new cultures of traceability articulate the phenomenon not only as a question of innovative methodologies, but also as an ethical concern. On the one hand, social scientists are understandably excited about the prospect of gaining more insights into the social dimension through access to the by-products of online human behaviour on a data set scale that matches that of the natural sciences (Venturini et al., 2015). On the other hand, historical experiences of surveillance and recent information scandals have also emphasised the volatile nature of digital traces and our vulnerability to them. Ironically, Cambridge Analytica’s Michael Kosinski warned about the dangers of digital traces as early as 2012:

importantly, given the ever-increasing amount of digital traces people leave behind, it becomes difficult for individuals to control which of their attributes are being revealed. For example, merely avoiding explicitly homosexual content may be insufficient to prevent others from discovering one’s sexual orientation. (Kosinski et al., 2013: 4)

While the entire spectrum of researchers, from protectors to exploiters of privacy, agree on the potential harm posed by digital traces, few scholars provide

programmatic definitions of what they mean by ‘digital traces’, instead letting the notion work as a loose heuristic for ‘fragments of past interactions or activities’ (Reigeluth, 2014: 250; see also Milan, 2018: 508). American sociologist Matthew J Salganik (2017: 71), for instance, treats digital traces as the by-product of people’s everyday actions, simply noting that ‘I’ve used the term of *digital traces*, which I think is relatively neutral’ compared with similar terms such as digital footprints and digital fingerprints. French sociologist Franck Cochoy et al. (2017: 4) use the notion in a similarly loose fashion, suggesting that these traces are what distinguish today’s datafied environment from previous information cultures: ‘chances are great that a main contribution of the digital world to contemporary society is the amazing proliferation of a new kind of entity: digital traces’. New media theorist Tyler Reigeluth (2014: 250) usefully adds a temporal dimension, reminding us that ‘digital traces are fragments of past interactions or activities which, when correlated together, allow a preemption and prediction of future behaviors’. Organisation theorists Mikkel Flyverbom and John Murray (2018) integrate this temporal aspect into an infrastructural problematic where digital traces are embedded in ‘datastructures’ that organise and order material ‘in ways that allow for analysis, value extraction and connection to different forms of social activity such as commercial production or political advocacy’. What make digital traces valuable, then, are the infrastructures that subtend the processing of data points, as well as the retemporalisation of these traces from past to future, making data actionable in new ways.

The Sciences Po Médialab ventures a more specific understanding of digital traces by distinguishing them from inscriptions and data: ‘by “digital traces”, we intend loosely all the inscriptions produced by digital devices in their mediation of collective actions – for instance, a post published on a blog, a hyperlink connecting two websites or the log of an e-commerce transaction’ (Venturini et al., 2017: 2). With the term ‘traces’, Venturini et al. (2017) thus refer to ‘inscriptions as originally produced by digital devices’ vis-à-vis data, which they use to refer to the ‘same inscriptions having undergone the cleaning and refining necessary to make them useful knowledge objects’ (3). While this distinction is more detailed than that provided by many other researchers, it is also more uncertain. Thus, as the authors concede themselves, their distinction ‘is somewhat artificial (there are no such things as “raw traces” and all inscription processes entail adjustments and correction)’ (Venturini et al., 2017: 3).

As the next section shows, this conceptual uncertainty about digital traces is not a bug, but a feature that emphasises the links between the politics of datafication and the politics of waste.

## Digital traces as by-products

In an article on digital traces, Deborah Maron of the University of North Carolina and Erin Carter of Cisco Systems note: ‘digital traces are an unavoidable byproduct of computer-mediated human interaction’ (Maron and Carter, 2017: 7). Several other data discourses echo this framing of digital traces as by-products (Jungherr, 2017; see also Giles, 2012; Howison et al., 2011; and many more).

This, then, is essentially what the Big Data hype is about: the waste-related epiphany that seemingly useless data can be extracted, recycled and resold for large amounts of money. An article in *Harvard Business Review* offers a good example of this sentiment: ‘today, companies in almost every industry are generating another valuable byproduct: data. Seemingly mundane accounting systems and customer databases now yield the raw materials that can be transformed into lucrative new services’ (Lewis and McKone, 2016). Such discourses on digital traces echo slogans such as ‘from trash to treasure’, and draw on the coding of trash as a resource and a rich arsenal of invention histories, where by-products suddenly become the main product (ranging, for example, from urine in toothpaste whitener, cow intestines in tennis rackets, coal tar in saccharine and swim bladders for wine-brewing to less fanciful by-products, such as bran and spirulina).

Viktor Mayer-Schönberger and Kenneth Cukier (2014: 112) provide us with a useful example of how value is extracted from digital waste. By analysing its users’ failed search attempts and typos, Google developed a multilingual and constantly updating spell-checker. While spellcheckers abound online, Google’s novelty was that it relied not only on correct data, but also on ‘defective’ data to create value. In the words of Mayer-Schönberger and Cukier (2014: 113), ‘only Google recognized that the detritus of user interactions was actually gold dust that could be gathered up and forged into a shiny ingot’.

Without linking to waste theories, new media theorist José Van Dijck (2014) places this form of rhetoric squarely at the heart of digital capitalism, noting that tech companies turned social activities into algorithmic relations. The latter in turn were ‘made accessible to third parties’, thereby spawning an ‘industry that builds its prowess on the value of data and metadata [...] not too long ago considered worthless byproducts of platform-mediated services’ but now ‘turned into treasured resources that can ostensibly be mined, enriched, and repurposed into precious products’ (Van Dijck, 2014: 199). Mayer-Schönberger and Cukier (2014: 113) echo this analysis and link it to the concept of data exhaust as ‘data that is shed as a

byproduct of people’s actions and movements in the world’, which companies can ‘harvest’ and ‘recycle’ to improve and innovate their products. But what does it mean to describe data traces as by-products, and what political ecologies does such a discourse yield?

## Pulverising and remoulding data traces

If memory traces are ‘signs in which remembering and forgetting are inextricably encoded’ (Assmann, 1996: 132), Big Data processes could be likened to the waste management processes that pulverise traces, turning them into a pulp from which new values can be sifted and mined (Bertolini, 1992). Such processes enable ‘a wide range of data about users and their digital traces to be folded back into data, in an endless cycle of “informating” happening inside digital infrastructures’ (Flyverbom and Murray, 2018).

A crucial component in this waste handling process is the divestment of identity from data. In order to divest the data they use of the clinging identities of the bodies from whence it came, Big Data companies need to subject it to a process ‘of pulverizing, dissolving and rotting’ to the point where ‘all identity is gone’, where ‘the origin of the various bits and pieces is lost’ and ‘they have entered into the mass of common rubbish’ (Douglas, 2001: 161). Removing identities from digital traces makes them less dangerous, since ‘where there is no differentiation there is no defilement’ (Douglas, 2001: 161). This process is less a question of finding fictitious ‘raw’ data, then, and more a question of ‘processing’ data in the right way so that it yields new insights without compromising the bodies from which it originated. As a report on open data and privacy states:

data needs treatment prior to publication. Often this involves “cleaning up” the data, removing egregious errors or inconsistencies, and generally improving quality. Preparation is sometimes needed to reduce the risk of the publication, by removing more sensitive aspects (from a privacy or commercial perspective) or by anonymising it such that it no longer constitutes personal data. (Simperl et al., 2016: 16)

This perspective shifts attention away from the oxymoron of ‘raw data’ (Gitelman, 2013), towards the complicated and material processes involved in treating data and turning it into recyclable resources for new forms of knowledge production (Boellstorff, 2013). With datafication we thus find ourselves in the waste incinerator, not the gold mine.

At this stage the next phase of data treatment begins. From their pulped state, digital traces are once again reinserted into the digital economy. As Walter Moser

(2002: 96) notes in his article on the acculturation of waste, this reintegration 'is possible only on the condition that things undergo a moment of negation as useful objects [...] that can assume various concrete forms: an act of rejection, total devaluation, or material destruction'. This is why the Big Data treatment is necessary: if data is to be presented as valuable by-products, it first has to be framed as initially useless and granular objects. The data companies' economic revaluation of waste therefore occurs 'less in the form of recovered or salvaged objects than as a formless mass that must undergo a process of recycling in order to once again become material' (Moser, 2002: 96). This negation is an important stage in the socio-economic process of data-structuring: if our activities are valuable as Big Data, they are only made so through the algorithmic resolution of digital traces into a data pulp that is then reassembled as valuable information in the form of new data, often going by the name 'data exhaust' or 'tertiary data' (Williams, 2013). Initially, these terms were used in the sense of 'by-product', signalling a new, revolutionary and potentially empowering use of digital traces.

In light of the frequent use of this notion of the by-product, it is relevant to take a closer look at its conceptual-legal framework and distinctive politics. According to the European Union's (EU) Directive 2008/98/EC on waste, a by-product is 'a substance or object, resulting from a production process, the primary aim of which is not the production of that item', and it can 'come from a wide range of business sectors, and can have very different environmental impacts' (European Commission, 2016). Drawing a clear distinction between waste and by-products thus offers us a semantic challenge. In a communication on how the EU's Waste Framework Directive should be interpreted, the European Commission (2007) states: 'in EU waste law, notions such as by-product or secondary raw material have no legal meaning – materials are simply waste or not'. In reality, however, there is 'not a black and white distinction, but rather a wide variety of technical situations with widely differing environmental risks and impacts and a number of grey zones' when one is deciding whether something is a by-product or pure waste, which is why definitions can only be made on a case-by-case basis (European Commission, 2007). This semantic instability echoes the classificatory issues raised by waste more broadly, and foregrounds the semiotic activity at work in waste (Douglas, 2001: 36). The relativity of the term 'waste' is dependent on a semiotic system of classification that has the power to determine one thing as waste, another as by-product, and yet another as main product. Hence, defining digital traces as by-products situates them in a legal grey zone, and also creates a cultural imaginary in

which digital traces are accidental rather than purposeful products. This grey zone opens up to the political decision-making power of whoever gets to determine the values and categories of waste, and when something is waste or value.

The following sections explore the politics of this data-recycling, and the ways in which the power to determine whether something is waste or value is unevenly distributed.

### Extracting value from toxic data sets

Machine learning relies on massive data sets where the data used for one purpose is repurposed for another. Facial recognition technologies pose particular analytical challenges, requiring enormous amounts of input to yield results. Offering facial recognition technologies the opportunity to validate their results, the United States' National Institute of Standards and Technology (NIST) launched a new facial recognition-testing programme in 2017. The purpose is to 'assess facial recognition systems on an on-going basis', and the programme will focus on how the tested systems perform with respect to 'accuracy, speed, storage and memory consumption, and resilience' (NIST, 2019). The basis of these tests is a data set of millions of images, which were collected for a different purpose but are now being used to test the algorithms. NIST is thus a classic example of Big Data-recycling: using digital by-products to create new inventions. It is also, as I shall argue, a perfect example of why the aforementioned semantic and legal grey zone of by-products raises the question not only of value but also of politics.

Although the data sets used to train facial recognition technologies are often treated as unremarkable, Os Keyes et al. (2019) recently published an article in *Slate* showing that the training data used by NIST contained images of people in vulnerable situations. They revealed that the United States' government's Facial Recognition Verification Testing programme depends on

images of children who have been exploited for child pornography; U.S. visa applicants, especially those from Mexico; and people who have been arrested and are now deceased. Additional images are drawn from the Department of Homeland Security documentation of travelers boarding aircraft in the U.S. and individuals booked on suspicion of criminal activity. (Keyes et al., 2019)

The management of this recycled product thus raises three fundamental questions regarding the data set's capacities for harm. First, the example reveals the

intimate link that remains between an image's 'by-product' and its provenance, even after the data has been 'processed' and 'pulverised' as Big Data. As the NIST report notes, the images NIST relies on in its data set 'are operational' and not meant to be read by human eyes (Grother et al., 2019: 18). However, as the founder of forensic science, Edmond Locard, once famously stated, 'every contact leaves a trace', and this applies to both the recovery of identities in data on a software level and the recovery of deleted information on a hardware level. In their report on digital forensics, Matthew Kirschenbaum et al. (2010: 45) paraphrase an expert's diagnosis of secure deletion as 'a major exercise' that 'can only be part of a secure "wipe" of one's entire hard disk. Anything less than that is likely to leave discoverable electronic evidence behind' (see also Caloyannides, 2001: 28).

Indeed, what is most valuable to Big Data companies is often precisely the capacity to connect a trace to its origin, at least performatively. For example, the promotional material for FindFace, a product by nTech, reveals that the company leverages its archive of images to offer extreme precision in its analytical work: 'FindFace Public Safety is able to simultaneously analyse data from hundreds of thousands of surveillance cameras, instantly distinguishing and storing the meaningful information (people's faces) from video stream' (nTech, n.d.). Of course, the linking of digital traces to identities relies more on a certain type of discourse than on an ontological reality, since the identity of a given object or subject is always already unstable. Hence, as Jacques Derrida (1976: 74) reminds us, 'a meditation upon the trace should undoubtedly teach us that there is no origin, that is to say simple origin; that the questions of origin carry with them a metaphysics of presence'. Few data companies, however, draw on Derrida's deconstructionist framework when proclaiming their own capabilities. Yet the performative effects of these companies' knowledge production fold identities into complex fabrications of truth in which the link between data and bodies persists.

In this way, the ghostly presence of those marked by violence ends up haunting Big Data's knowledge production processes through their digital traces, just as the knowledge produced by datafication in turn comes back to haunt those marked by structural inequalities (Blackman, 2020). This insight points to the need to reinforce public debates and establish legislative frameworks concerning the ethics of reuse. Researchers on colonial imagery have long raised concerns about the ethics of recycling, recirculating, and repurposing colonial images (Agostinho, 2019; Danbolt, 2017; Johnson, 2018; Meyer, 2016;

Sutherland, 2020). These concerns can meaningfully be mobilised in conjunction with discard studies' nuanced perspective on the material processes of recycling, to challenge the apparent truism that data is 'just data', and to amplify the assertions already made within computational and social justice communities that reuse practices such as those of NIST disregard the residual presence of humans in their data sets (Amoore, 2019; Keyes et al., 2019). Indeed, the NIST example shows that traces of vulnerable individuals haunt not only the standards developed by NIST, but also all the facial recognition technologies that are tested against those standards, which are thereby entangled within larger structural inequalities. As Jacqueline Wernimont noted in a conversation with me recently, the government's handling of the NIST case resembled the green washing of toxic datasets as good and responsible 'recycling'. At a deeper level, then, the NIST example also underlines how Big Data ecologies rely 'on disjunctures and contradictions (for discursive/greenwashing purposes in particular)' (Hogan, 2019): tech companies increasingly 'partner with/enslave' people 'in order to maintain and grow [their own] operations', while also demonstrating concern for those same people via large-scale infrastructural developments deployed to care for vulnerable subjects (Hogan, 2019).

## Concluding remarks

Digital traces are at the crux of Big Data knowledge production. While often used to describe large amounts of data, the notion of the digital trace has also come to connote that these Big Data sets are haunted and can haunt us in return (Blackman, 2020). Moreover, the material quality of the trace reminds us of 'the hazards posed by the massive computation of data on an increasingly fragile environment' (Gregg, 2015: 46), and of how our mediated existence hovers in a complex and toxic techno-ecology underpinned by the polluted work of mining for minerals and 'disposing of toxic creations of our own making' (Hogan, 2015). The knowledge production of Big Data thus appears before our eyes not only as a computational phenomenon that can yield new insights, but also as a recent point in a much longer history of production and destruction, power and suppression. As Aleida Assmann (1996: 132) reminds us, Jonathan Swift already exposed the tendency towards restless and reckless innovation in relation to print media, describing it as a dialectical process of production and rubbishing. Digital traces belong to this dialectical process, making datafication as much a material problematic of waste

management as an analytical problem of information. As the NIST example shows, the knowledge production processes of Big Data thus present us with a new form of waste colonialism that extends beyond the extraction of minerals and the disposal of mineral e-waste (Hogan, 2019; Parikka, 2015) and into racialised, gendered and classed data sets. This form of value extraction from data-recycling reproduces a biopolitics of disposability (Mbembe, 2016) where individuals become by-products of the data capitalist mode of knowledge production.

### Acknowledgements

I gratefully acknowledge the support and generosity of Danish Research Council's YDUN programme, without which the present article could not have been completed and Mer Storr's wonderful and always attentive proof-reading, without which the article would not have been readable. I would also like to express my gratitude to Kristin Veel and Daniela Agostinho, who have read and offered invaluable perspectives to previous versions of this article. Moreover, I thank my co-speakers on the panel on 'Digital Excess' at ICA 2019, including Fenwick McKelvey, Nora Draper, Elizabeth Wissinger and Stephanie Schulte, and our audience for enriching feedback. Finally, I extend a warm thank you to the anonymous peer reviewers, who have generously and in a constructive manner given their time and shared their knowledge to help me sharpen my arguments.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Nanna Bonde Thylstrup  <https://orcid.org/0000-0001-6094-2970>

### References

- Agostinho D (2019) Archival encounters: Rethinking access and care in digital colonial archives. *Archival Science* 19(2): 141–165.
- Agre PE (1994) Surveillance and capture: Two models of privacy. *Information Society* 10(2): 101–127.
- Alexander VD, Blank G and Hale SA (2018) Digital traces of distinction? Popular orientation and user-engagement with status hierarchies in TripAdvisor reviews of cultural organizations. *New Media & Society* 20: 4218–4236.
- Amoore L (2019) Cloud ethics. Presentation given at Copenhagen Business School, 23 April.
- Assmann A (1996) Texts, traces, trash: The changing media of cultural memory. *Representations* 56: 123–134.
- Bertolini G (1992) Les déchets: Rebut ou ressources? *Économie et statistique* 258–259: 129–134.
- Bigo D (2006) Security, exception, ban and surveillance. In: Lyon D (ed.) *Theorizing Surveillance*. New York, NY: Routledge, pp. 46–68.
- Blackman L (2020) Hauntology. In: Thylstrup N, Agostinho D, Ring A, et al. (eds) *Uncertain Archives*. Cambridge, MA: MIT Press.
- Boellstorff T (2013) Making big data, in theory. *First Monday* 18(10). Available at: [firstmonday.org/ojs/index.php/fm/article/view/4869/3750%232](http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750%232) (accessed 1 July 2019).
- Boullier D (2017) Big data challenges for social sciences: From society and opinion to replications. *ISA eSymposium* 7(2). Available at: [www.sagepub.net/isa/admin/ebulletin-articles.aspx](http://www.sagepub.net/isa/admin/ebulletin-articles.aspx) (accessed 1 July 2019).
- Caloyannides M (2001) *Computer Forensics and Privacy*. Norwood, MA: Artech House.
- Chun WHK (2016a) Big data as drama. *EHL* 83(2): 363–382.
- Chun WHK (2016b) *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press.
- Cochoy F, Hagberg J, Petersson M, et al. (2017) *Digitalizing Consumption: How Devices Shape Consumer Culture*. Abingdon: Routledge.
- Cohen JE (2012) *Configuring the Networked Self: Law, Code, and the Play of Everyday Practices*. New Haven, CT: Yale University Press.
- Cubitt S (2014) Decolonizing ecomedia. *Cultural Politics* 10(3): 275–286.
- Danbolt M (2017) Retro racism: Colonial ignorance and racialized affective consumption in Danish public culture. *Nordic Journal of Migration Research* 7: 105–113.
- Deleuze G (1992) Postscript on control societies. *October* 59: 3–7.
- Derrida J (1976) *Of Grammatology*. Baltimore, MD: John Hopkins University Press.
- Douglas M (2001) *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London/New York, NY: Routledge.
- European Commission (2007) Communication from the Commission to the Council and the European Parliament on the interpretative communication on waste and by-products. Available at: [eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52007DC0059&from=EN](http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52007DC0059&from=EN) (accessed 1 July 2019).
- European Commission (2016) Directive 2008/98/EC on waste. Available at: <http://ec.europa.eu/environment/waste/framework/> (Accessed April 15, 2019).
- Flyverbom M and Murray J (2018) Datastructuring – Organizing and curating digital traces into action. *Big Data & Society* 5(2). <https://journals.sagepub.com/doi/full/10.1177/2053951718799114>
- Gabrys J (2010) *Digital Rubbish: A Natural History of Electronics*. Ann Arbor: University of Michigan Press.
- Gabrys J (2016) *Program Earth: Environmental Sensing Technology and the Making of a Computational Planet*. Minneapolis: University of Minnesota Press.
- Giles J (2012) Making the links. *Nature* 488(23): 448–450.



- Gitelman L (2013) *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press.
- Gregg M (2015) Inside the data spectacle. *Television & New Media* 16(1): 37–51.
- Grother P, Ngan M and Hanaoka K (2019) Ongoing face recognition vendor test (FRVT): Part 1: verification, April 4. Available at: [https://www.nist.gov/sites/default/files/documents/2019/04/04/frvt\\_report\\_2019\\_04\\_04.pdf](https://www.nist.gov/sites/default/files/documents/2019/04/04/frvt_report_2019_04_04.pdf) (accessed 26 April 2019).
- Hall S (2013) *Representation: Cultural Representations and Signifying Practices*. London: Sage.
- Hepp A, Breiter A and Friemel TN (2018) Digital traces in context: An introduction. *International Journal of Communication* 12: 439–449.
- Hesselberth P (2018) Discourses on disconnectivity and the right to disconnect. *New Media & Society* 20(5): 1994–2010.
- Hogan M (2015) Facebook data storage centers as the archive's underbelly. *Television & New Media* 16(1): 3–18.
- Hogan M (2019) Big data ecologies. *Ephemera: Theory & Politics in Organization* 18(3). Available at: [ephemerajournal.org/contribution/big-data-ecologies](http://ephemerajournal.org/contribution/big-data-ecologies) (accessed 26 April 2019).
- Howison J, Wiggins A and Crowston K (2011) Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems* 12(12): 767–797.
- Johnson JM (2018) Markup bodies. *Social Text* 36: 57–79.
- Jungherr A (2017) Normalizing digital trace data. In: Jomini Stroud N and McGregor S (eds) *Digital Discussions: How Big Data Informs Political Communication*. New York, NY: Routledge, pp. 9–35.
- Keyes O, Steven N and Wernimont J (2019) The government is using the most vulnerable people to test facial recognition software. *Slate Magazine*, 17 March. Available at: [slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html](http://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html) (accessed 29 April 2019).
- Kirschenbaum MG, Ovenden R, Redwine G, et al. (2010) *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington, DC: Council on Library and Information Resources.
- Kitchin R (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1. Epub ahead of print 2014. Available at: <https://journals.sagepub.com/doi/abs/10.1177/2053951714528481>.
- Kosinski M, Stillwell D and Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110(15): 5802–5805.
- Lepawsky J (2018) *Reassembling rubbish worlding electronic waste*. Cambridge, MA: MIT Press.
- Lewis A and McKone D (2016) To get more value from your data, sell it. *Harvard Business Review*, 21 October. Available at: [hbr.org/2016/10/to-get-more-value-from-your-data-sell-it](http://hbr.org/2016/10/to-get-more-value-from-your-data-sell-it) (accessed 1 July 2019).
- Liboiron M (2016) Redefining pollution and action: The matter of plastics. *Journal of Material Culture* 21(1): 87–110.
- Liboiron M, Tironi M and Calvillo N (2018) Toxic politics: acting in a permanently polluted world. *Social Studies of Science* 48(3): 331–349.
- Lyon D (2008) *Surveillance Studies: An Overview*. Cambridge: Polity Press.
- Mai J-E (2016) Big data privacy: The datafication of personal information. *Information Society* 32(3): 192–199.
- Maron D and Carter E (2017) 'More than what it seems': How critical theory, popular engagement and apps like Tinder can help us reframe metadata and its consequences. In: *Proceedings of the international conference on Dublin core and metadata applications*. Dublin Core Metadata Initiative (DCMI), A project of ASIS&T, Washington, DC, April 15, 2019, pp. 1–12. Available at: <http://dcevents.dublincore.org/public/dc-docs/2017-Master.pdf>
- Mayer-Schönberger V and Cukier K (2014) *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight*. London: John Murray.
- Mbembe A (2016) The society of enmity. *Radical Philosophy*. Available at: <https://www.radicalphilosophy.com/article/the-society-of-enmity>
- Meyer MKK (2016) Blind Spots. Lecture given at Danish Royal Library, Copenhagen, 5 November.
- Milan S (2018) Political agency, digital traces, and bottom-up data practices. *International Journal of Communication* 12: 507–527.
- Moser W (2002) The acculturation of waste. In: Neville B and Villeneuve J (eds) *Waste-Site Stories: The Recycling of Memory*. Albany: State University of New York Press, pp. 85–106.
- National Institute of Standards and Technology (NIST) (2019) FRVT 1:1 verification. Available at: [www.nist.gov/programs-projects/frvt-11-verification](http://www.nist.gov/programs-projects/frvt-11-verification) (accessed 29 April 2019).
- Nissenbaum H (2010) *Privacy in Context: Technology, Policy and the Integrity of Social Life*. Stanford, CA: Stanford Law Books.
- nTech (n.d.) FindFace Public Safety. Available at: [findface.pro/en/face-recognition-public-safety.html](http://findface.pro/en/face-recognition-public-safety.html) (accessed 29 April 2019).
- Parikka J (2011) New materialism as media theory: Media natures and dirty matter. *Communication and Critical/Cultural Studies* 9(1): 95–100.
- Parikka J (2015) *A Geology of Media*. Minneapolis: Minnesota University Press.
- Räsänen M and Nyce JM (2013) The raw is cooked: Data in intelligence practice. *Science Technology and Human Values* 38(5): 655–677.
- Reigeluth T (2014) Why data is not enough: Digital traces as control of self and self-control. *Surveillance and Society* 12(2): 243–254.
- Ruppert E, Isin E and Bigo D (2017) Data politics. *Big Data & Society* 4(2). Available at: <https://journals.sagepub.com/doi/abs/10.1177/2053951717717749>.
- Salganik MJ (2017) *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.
- Simperl E, O'Hara K and Gomer R (2016) Analytical report 3: Open data and privacy. Available at: [www.europeandataportal.eu/sites/default/files/open\\_data\\_and\\_privacy\\_v1\\_final\\_clean.pdf](http://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf) (accessed 29 April 2019).

- Sutherland T (2020) Remains. In: Thylstrup N, Agostinho D, Ring A, et al. (eds) *Uncertain Archives*. Cambridge, MA: MIT Press.
- Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Venturini T, Bounegru L, Gray J, et al. (2018) A reality check(list) for digital methods. *New Media & Society* 20(11): 4195–4217. <https://journals.sagepub.com/doi/abs/10.1177/1461444818769236>
- Venturini T, Jacomy M, Meunier A, et al. (2017) An unexpected journey: A few lessons from Sciences Po Médialab's experience. *Big Data & Society* July–December. Available at: [journals.sagepub.com/doi/pdf/10.1177/2053951717720949](https://journals.sagepub.com/doi/pdf/10.1177/2053951717720949) (accessed 1 July 2019).
- Venturini T, Jensen P and Latour B (2015) Fill in the gap: A new alliance for social and natural sciences. *Jasss* 18(2): 1–4.
- Welser HT, Smith M, Fisher D, et al. (2010) Distilling digital traces: Computational social science approaches to studying the internet. In: Fielding NG, Lee RM and Blank G (eds) *The Sage Handbook of Online Research Methods*. Los Angeles, CA: Sage, pp. 116–140.
- West SM (2017) Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society* 58(1): 20–41.
- Williams A (2013) The power of data exhaust. *TechCrunch*, 26 May. Available at: [techcrunch.com/2013/05/26/the-power-of-data-exhaust/](http://techcrunch.com/2013/05/26/the-power-of-data-exhaust/) (accessed 1 July 2019).
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.