



Your data is s**t*

Robert Herian

abstract

Personal data is the chief commercial and informatic industrial raw material of the last forty years. But as an almost universal daily excretion composed of body with environment (personal with non-personal data), so much personal data is shit. This note addresses a data metaphor that seeks to explain and situate personal data and the data subject socially, politically, economically, and legally. Data models do not want messiness (shit) or inefficiency, only simple and logical input/output risk defiant certainties concerning population types and cohorts. But tending to the growing hot heaps of data involves an expanding complex of systems, networks, frameworks, rules, mechanisms, policies, and ideologies of governance and governmentality, both on- and offline. I call this complex a shitshow. Strategies for individuals, organizations, and economies are of paramount interest and concern as each attempt to navigate the shitshow. Echoing the work of Dominique Laporte, I consider how the shitshow leads to data hygiene practices for managing storage, cleansing, and refinement of shit data, and, increasingly, to extract profit from it.

* I am grateful and indebted to the reviewer of this note for their excellent feedback. I would also like to thank friends, colleagues, participants and panellists at the Critical Legal Conference, University of Dundee, September 2021, and the Law, Technology and the Human Conference, University of Kent, April 2022, for their important comments and insights on earlier versions of this note.

Welcome to the shitshow

In examining new ways for understanding the contemporary data subject, I believe the metaphor of shit is useful. It reminds us that data is a daily excretion composed of body with environment (personal with non-personal data). Like sludge (refined sewage), data excretions provoke, what I refer to here as, hygiene processes to manage storage, cleansing, and refinement, and increasingly realise inherent value (Hope, 2016). As a result, like common or garden compost heaps, data servers radiate more heat as the storage of petabytes of personal data ‘piles-up’.¹

As a seemingly unargumentative source of value, personal data has become the chief commercial and informatic industrial raw material of the last forty years. An asset class par excellence, personal data proliferation due to rapidly increasing levels of computer use (including, notably, mobile devices) has been a boon in recent years for domestic and international data brokage (see, for example, Sherman, 2021). I want to continue the discussion I started in *Data: New trajectories in law* (Herian, 2021) on data metaphors that help explain and situate personal data and the data subject socially, politically, economically, and legally.

Today, tending to the growing hot heaps of data involves an expanding complex of systems, networks, frameworks, rules, mechanisms, policies, and

¹ I recognise linkages to Donna Haraway’s (2016) use of the term ‘compost’ here, or at least a benefit that may derive to my refinement of the concept of shit data from reading it back through Haraway’s work. Haraway’s own preference for the notion of compost over or in critical contradistinction to the term ‘post-human’ is certainly noteworthy. As Haraway mentions in an interview with Sarah Franklin (2017: 51-52):

I like the word "compost" because it includes living and dying. If you're in compost, the questions of finitude and mortality are prominent, not in some kind of depressive or tragic way, but those who will return our flesh to the Earth are in the making of compost. I can't work my compost pile without being in the midst of the question of how to inherit the multiple histories and the multiple formations that allow this compost pile to be cooking badly in my yard, you know. They are provocations to becoming more historical, in the sense of bringing what you inherit into the present so as to somehow become more able to respond.

ideologies of governance and governmentality, both on- and offline. Spanning commercial and non-commercial sectors, the hot heaps of data excite, enthrall, occupy, and burden private and public bodies and individuals (i.e., data subjects) simultaneously ignorant and interpolated in vertical and horizontal domains of organization. Despite legislative and regulatory interventions, notably but not only Europe's General Data Protection Regulation (GDPR), it is not always clear where boundaries of authority and responsibility lie regarding the giving and receiving of personal data or its subsequent conveyance, use, and exploitation. Therefore, we might rightfully and, on terms I rely on here, also accurately call the situation data subjects find themselves in today a shitshow.

My previous interest in rethinking data was to understand data autonomy as data which excuses, alludes, or exceeds human need, demand, and desire. Data without need of a subject, and, we might argue, therefore utterly without use, value, or purpose where it cannot register either in human perception or via the tools and technologies built to enhance human perception. Data models do not want messiness (the shit) or inefficiency, only simple and logical input/output risk defiant certainties concerning population types and cohorts. Hence, corresponding data rhetoric and narratives able to explain to individuals, organizations, and economies more broadly, the incontestable value of data are of paramount interest and concern.

Rhetoric of techno hygiene

Your data is shit: this expression contains many ways of understanding humanity's relationship with petabytes of data produced in the present technological moment. For instance, your data is shit because you, as an individual, provide little or no value to medical science despite the constant streams of data produced by your wearable tech; your data is shit because you, as an entrepreneur, cannot leverage insights for maximum commercial benefit from the app you built and the data it captures; your data is shit because you, as a corporation, have failed to see profitable returns for shareholders on a series of advertising campaigns for your latest product. These interpretations speak to data's value rooted squarely in a discourse of innovation and progress. More than that, however, we must understand data

narratives as a product of neoliberal stakeholders and the markets they aim to birth (or leverage) at every opportunity. Shit data is seemingly of little use or obvious profit, yet commercial and non-commercial stakeholders routinely gather and keep it, often with a feverish endeavour.

Techno-hygienists today surveil and collate humanity's mass digital excretions and extrusions, capturing them more pervasively and with ever greater levels of sensitivity, machinic power and sophistication. Treatment and processing of informatic ordure along with techniques of purification, filters out value. This is important because, as Cox et al. (2012: 75) explain, 'like the tradition of examining feces to determine the health of the organism [a practice given additional urgency during the Covid-19 pandemic], the health of the economy can be judged by the way it manages its waste.'

Describing the hygienic revolution undertaken over several centuries across Western capitalist societies, Laporte (2002: 118-119) considers the perception of the hygienist as a hero when it was 'no longer enough to eliminate and separate shit into solid and liquid components, to flush and disinfect it. [S]hit', Laporte argues, had 'to become profitable' The hygienists achieved this end, the realising of value from shit, with heroic endeavour.

Today, we find this continuing rhetoric of hygiene enables markets around technologies for and techniques of data self-care, prompting unending rituals, practices, and performances of data hygiene that construe every individual a hero worthy of endowment and reward when they manage data effectively, efficiently, and profitably. Importantly, information capitalism increasingly promotes a role or perhaps even an ethical duty for consumers, as data subjects, to take control (and ownership) of 'their' data, to 'get their shit together' so to speak and monetize it whenever and wherever possible, notably by submitting to tailor-made advertising (see, for example, <https://gener8ads.com/>).

'Some shit is incontestably good,' Laporte (2002: 111) claims,

... not just because it has been purified, but because it is that which purifies. It purifies because it is spirit and soul – a volatilization of the flesh that retains an attachment to the body from which it has been severed. Shit never stops being a fragment of God.

Questions of the extent of the retention of data from the ‘body from which has been severed’ underpin much of the developing contemporary regulatory and legislative emphasis on privacy, data, and consumer rights and protections. But these legal interventions have not stopped the flow. There is no sign of data constipation among global populations. Quite the opposite. The known global internet population continues to grow year on year to over 4.5 billion in 2020 and streams of data flowing into what Julie Cohen (2019) calls the biopolitical public domain intensifies. YouTube boasts the addition of 500 hours of new content per minute, WhatsApp over 42 million messages in the same timeframe, to name just two predominate sites of normative data practice and performance today (www.domo.com, 2020). Also, legal frameworks, or regulatory reluctance to interfere with innovation, ultimately support intensification of data flows. ‘The data flows extracted from people play an increasingly important role as *raw material* in the political economy of informational capitalism’, argues Cohen (2019: 48). Continuing,

... personal data processing has become the newest form of bioprospecting, as entities of all sizes – including most notably both platforms and businesses known as data brokers – compete to discover new patterns and extract their marketplace value. Understood as processes of resource extraction, the activities of collecting and processing personal data require an enabling legal construct. (*ibid.*: 48)

Cultivated and extracted data enter an industrial production process during which they are refined to generate data doubles – information templates for generating patterns and predictions that can be used to optimize both online and physical environments around desired patterns of attention and behaviour [...] the participants in the data economy trade in people the way one might trade in commodity or currency futures. (*ibid.*: 64)

Producing data/shit

As the amount of shit produced by internet users increases, the so-called ‘market for eyeballs’ thrives, underscored by internet platform business models reliant on capturing and extending user attention and engagement on behalf of advertisers.² These models are far from ephemeral. Instead, each

² Personal financial data has more of a role to play in such models. For instance, the possible advent of Central bank Digital Currencies (CBDCs), cryptocurrency aimed at supplementing or superseding fiat money, notably cash, will probably end up

relies on high levels of data input, through-flow, and storage to prevent loss and maximise benefit for businesses over the medium and long term. ‘For the hygienists’, Laporte (2002: 124) suggests,

... shit was the site of irredeemable, even incommensurable loss, which they were obstinately bent on denying. They were caught in a tenacious thwarting of loss that sustained their delirious claim to matter, their heroic compulsion to retain. Their discourse, although synchronous with capitalism, is not the discourse of capitalism, but its symptom.

Again, despite constraints created by the likes of GDPR in Europe and California’s Consumer Privacy Act (CCPA), the compulsion for internet users to engage with platforms and open themselves to being sourced (as sources of data) remains strong. Platforms are the products of contemporary hygienists, designed to give users clear (if not always hospitable) social interfaces. As increasingly indispensable points of intermediation, platforms attract huge numbers of users and, as a result, harvest tremendous amounts of data, with approximately half the global population, 3.5 billion people, use social networks alone (www.statista.com, 2022).

All data is shit and to produce, as Laporte (2002: 131) says, ‘is literally to shit’. Global data storage adds to what the International Data Corporation (IDC) calls the ‘global DataSphere’ (www.blogs.idc.com, 2019). This seemingly unrestrained global data production is facilitated by sensors in billions of interconnected devices and filtered and processed by increasingly rapid forms of machine learning and automation. As a result, in today’s data rich environments – more than 79.4ZB of data created by 2025

linking with individual bank accounts rather than maintaining the anonymity many people favour with the various decentralized cryptocurrencies we see today. As Izabella Kaminska (2021) in the *Financial Times* suggests,

‘if money is to be identity-based rather than token-based and fungible, this introduces a whole new set of ethical dilemmas and social questions, which aren’t really being asked at the moment on a wide enough social level. The conversations we should be having relate to who do we as a society really entrust with our personal data? The current choice includes private companies like Facebook, highly regulated private institutions like banks, “independent” central banks, government-directed central banks, a bit of everyone or nobody at all.’

(www.blogs.idc.com, 2019) – data as a by- or waste product, spin-off, or data exhaust, and so-called ‘dark data’ are influential ideas that account for a desire and need for ensuring more and better commercial use and value from the excess, hot (composting) heaps of personal data.

Wasteful relations

Jane Bennett provides two important arguments for thinking about humanity’s relationship with waste products, of which we must now surely include data. The first concerns the force exerted by thingly-power as ‘vivid entities not entirely reducible to the contexts in which (human) subjects set them’ (Bennett, 2010: 5), and the second concerns the agency of things that ‘always depends on the collaboration, cooperation, or interactive interference of many bodies and forces’ (*ibid.*: 21). Bennett’s account of things exceeding humanity’s perception of or interest in them, or as Bennett (*ibid.*: 4) puts it, things ‘in excess of their association with human meanings, habits, or projects’, is key to understanding the ‘afterlife’ of things, discarded or used-up by humanity. This is, for Bennett, a sign not of where the being of things ends but where it arguably becomes most prominent, and its vitality begins. Things that humanity no longer has a use for or sensory interest in (to see, hear, smell, or touch, etc.), do not make them cease to exist in the world. Instead, they continue as their own particular and peculiar manifestation of non-organic life and being.

Human-made categorisations distinguishes between things once considered within human perception to be what we might call ‘useful’, and those things that don’t – what we routinely called ‘waste’, ‘junk’, or ‘refuse’ and may also add the concept of shit data to – as a source of meaning and reality. But it is not reality. It is quite the opposite, in fact: we predicate categorization solely on a guarantee of human perceptive authority and power, which is granted to humanity by itself. Hence, for Bennett (2010: 6) ‘a vital materiality can never really be thrown “away”, for it continues its activities even as a discarded or unwanted commodity’. This idea does, albeit tangentially, correspond with mathematician David Hand’s (2019) view of dark data as classifications of data given meaning by how we collect them.

The human conclusion as to and categorisation of waste (debris, trash, litter, etc.) is important to Bennett's (2010: 5) exposition of things as 'vibratory – at one moment disclosing themselves as dead stuff and at the next as live presence: junk, then claimant; inert matter, then live wire'. And, I suggest, this offers us a way to frame an understanding not only of human data production(s), but of the systematic and systemic ways in which to conceptualise and actualise production. Now, it seems, we were wrong to ignore shit data. 'If we are clever enough', argues David Hand (2019: 5), 'we can sometimes take advantage of dark data. Curious and paradoxical though that may seem, we can make use of ignorance and the dark data perspective to enable better decisions and take better actions. In practical terms,' Hand (2019: 5) concludes, 'this means we can lead healthier lives, make more money, and take lower risks by judicious use of the unknown'. Referring to data that describe humans as administrative data, able to 'tell you *what people do*' and 'get you nearer to social reality than exercises involving asking people what they did or how they behave', Hand (2019: 31) explains databases full of personal or administrative data 'represent a great resource, a veritable gold mine of potential value enabling all sorts of insights to be gained into human behaviour'.

The data subject is at once an individual bringer and giver of data and receiver of rights and protections of and over the stuff called 'personal data', where lawmakers attribute such data to them as set out within legislation. But the data subject is also one who is often in ignorance, one for whom the status and nature of personal data is at once mysterious and burdensome. Whilst increasingly intimately associated with technologies like smartphones and the technological know-how that accompanies them, data subjects vary in awareness as to their status as sources of personal data or understanding of its value or fate as it circulates within global capitalist economies. Data subject awareness of their productive value within informational capitalism, although arguably related to labour processes, differs from an assumption made about workers elsewhere in capitalism, that they have a good awareness of their working conditions and their exploitation within capitalism must, therefore, take indirect forms, notably pricing (Chibber, 2022).

Pricing is a method yet to act *en masse* against the beneficial interests of data subjects who are, by and large, still producing the volumes of shit techno-

hygienists crave on daily basis, and at very low cost. This concerns not only the data subject's reliance upon the intermediary services provided by platforms in the 'management' of the subject's data (the basis of the adage attributed to the American artist Richard Serra 'if something is free, you are the product'), but a powerful belief in the legitimacy of data sovereignty and of capturing a dimension of data labouring experience able to fend off a lack of personal discipline and the risk of squandering the value of the subject's shit data. 'Shit is productive only insofar as it is human', Laporte (2002: 120) reminds us, 'of all the other manures known to nature, none is equal to human fertilizer'. In personal data today, we find the productivity of human shit elevated to new and transcendental levels, body with environment, material in virtual.

The shitshow continues (a conclusion)

And this brings me back to those who seek to control personal data; to cultivate, extract, and exploit data systemically for value with increasing precision and sophistication – what I have referred to throughout as techno-hygienists, echoing Dominique Laporte's history of shit. Laporte (2002: 133) describes the terms upon which State norms established expectations on the subject to manage their shit, claiming that:

Shit is the precious object par excellence, the object that must not be squandered at any cost. But it is equally that which the subject must renounce, "religiously collect," and deliver to the State under a double burden: on one hand, the promise of an end to lack and, on the other, the threat of hardship, given a lack of discipline.

For me, Laporte could just as easily be talking about the bargain struck by the data subject not just with the State (of course, this bargain depends on the State in question), but with platforms and other stakeholders of the commercial Internet whose uniformity of purpose is arguably more clear-cut than the State: to know your shit.

And, finally, lest we forget the abundant associations between our on- and offline worlds, despite obvious gaps between the two, those slippages purveyors of the 'metaverse', would have us enjoy in the face of climate catastrophe, Brian Thill (2015: 26-27), like Jane Bennett, reminds us that

‘digital waste is not freed from the realities of material existence. Just like the coffee we drink, its ongoing production consumes immense energy, labor, resources, time, and space, just as all the proliferating garbage of the pre-digital ages did and continues to do.’

references

- Bennett, J. (2010) *Vibrant matter: A political ecology of things*. Durham: Duke University Press.
- Chibber, V. (2022) *The class matrix: Social theory after the cultural turn*. Cambridge: Harvard University Press.
- Cohen, J.E. (2019) *Between truth and power: The legal constructions of informational capitalism*. Oxford: Oxford University Press.
- Cox, G., A. McLean and F. Berardi (2012) *Speaking code: Coding as aesthetic and political expression*. Cambridge: MIT Press.
- Franklin, S. (2017) ‘Staying with the manifesto: An interview with Donna Haraway’, *Theory, Culture & Society*, 34(4): 49-63.
- Hand, D.J. (2019) *Dark data: Why what we don’t know matters*. Princeton: Princeton University Press.
- Haraway, D. (2016) *Staying with the trouble. Making kin in the chthulucene*. Durham: Duke University Press.
- Herian, R. (2021) *Data: New trajectories in law*. Abingdon: Routledge.
- Hope, K. (2016) ‘The firms turning poo into profit’, *BBC News* [<https://www.bbc.co.uk/news/business-37981485>].
- Kaminska, I. (2021) ‘Why CBDCs will likely be ID-based’, *Financial Times* [<https://www.ft.com/content/88f47c48-97fe-4df3-854e-0d404a3a5f9a>].
- Laporte, D. (2002) *History of shit*. Cambridge: MIT Press.
- Sherman, J. (2021) ‘Data brokers are a threat to democracy’. *Wired* [<https://www.wired.com/story/opinion-data-brokers-are-a-threat-to-democracy/>].
- Thill, B. (2015) *Waste*. London: Bloomsbury.

www.blogs.idc.com (2019) 'How you contribute to today's growing datasphere and its enterprise impact'. *IDC* [<https://blogs.idc.com/2019/11/04/how-you-contribute-to-todays-growing-datasphere-and-its-enterprise-impact/>]

www.domo.com (2020) 'Data never sleeps 8.0'. *Domo* [<https://www.domo.com/learn/infographic/data-never-sleeps-8>]

www.statista.com (2022) 'Most popular social networks worldwide as of January 2022, ranked by number of monthly users' [<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>]

the author

Dr Robert Herian's current research focuses on interdisciplinary and theoretical analyses of law, technologies, systems, and data. He works on law and technology policy development with UK and EU governments, has presented research at domestic and international conferences, and published in legal and non-legal peer-reviewed journals, edited collections, and via online portals including *The Conversation* and *Critical Legal Thinking*. Dr Herian is also author of three books *Regulating blockchain: Critical perspectives in law and technology* (Routledge, 2018), *Data: New trajectories in law* (Routledge, 2021), and *Capitalism and the equity fetish: Desire, property, justice* (Palgrave Macmillan, 2021).

Email: r.herian@exeter.ac.uk