

Critical data futures

Neil Selwyn
Monash University, Melbourne

**chapter for Housley, W., Edwards, A., Montagut, R. and Fitzgerald, R. (eds).
The Sage Handbook of Digital Society.
[second draft: 06_08_21]**

INTRODUCTION

The past decade has seen mainstream social science at last begin to turn attention toward the digital. At the same time, social scientists who have long been interested in digitally-related topics have shown increasing interest in issues relating to data. As Couldry and Mejias (2019, p.336) observe dryly, the digital social sciences are now ablaze with a sense that “something important is going on with data”. Of course, data in analogue form has long been a heavily scrutinized feature of modernity – not least in terms of statistics, record-keeping, standardized measurement, and other tenets of the bureaucratic state. More specifically, then, the digital social sciences have latched onto what is termed the ‘datafication’ of contemporary society. This relates to the masses of ‘trace’ data that are now generated through the use of digital technologies – often unbeknown to the people engaging with these technologies - as well as the processing of huge volumes of data by powerful computational systems.

The social significance of digital ‘data’ became increasingly difficult to ignore as the 2010s progressed – initially via corporate and policy enthusiasms for so-called ‘Big Data’, and then in the steady societal rise of algorithmic systems and various forms of data-driven decision making. These trends look set to continue throughout the 2020s – a decade that will see expected expansion of billions of data sensors (aka the ‘Internet of Things’), alongside the continued development of artificial intelligence technologies and other forms of data-based automations. As such, digital data is set to be a significant point of interest for digital social scientists over the next decade and beyond. To complement other chapters in this handbook concerned with various aspects of data and datafication (for example, Chapters X, Y, and Z), this particular chapter addresses the topic as part of the handbook’s ‘Digital Futures’ section. So, what does it mean to talk about ‘data futures’? More pertinently, what aspects of recent critical scholarship around the datafication of contemporary society might be useful when anticipating possible data futures?

This chapter progresses through three phases of discussion. First, we consider a number of different ‘lenses’ that emerge from recent social science work within the critical data studies tradition. Second, we reflect on the growing trend for futures thinking, imaginaries, anticipatory studies, and other futures-focused scholarship within the social sciences. Finally, we present initial outlines of two different anticipatory scenarios that expand on how these critical data ‘lenses’ might play out in future forms of data-driven society – each scenario foregrounding a number of

broader concerns relating to changing forms of ontology, ideology, power and political consequences. The chapter concludes by considering briefly how future scenarios of data-driven 'digital societies' might draw attention to unfolding reconfigurations of data regulation, data governance, data resistances and data responsibilities – all of which need to be foregrounded in terms of debates around the ongoing digitisation of societies and the everyday lives of those who live within them.

1. THE RISE OF 'CRITICAL DATA STUDIES'

As detailed elsewhere in this handbook (Chapters X, Y and Z), social science accounts of digital data have begun to increasingly coalesce around what has become known as 'critical data studies'. As with most contemporary scholarship relating to digital matters, critical data studies is an eclectic, interdisciplinary endeavour – perhaps best described as a loose 'research theme' rather than a tightly focused 'research field'. Indeed, Dalton *et al.* (2016, p.1) concede that critical data studies might be recalled in years to come as little more than "three words cobbled together imperfectly signifying diverse sets of work around data's recursive relationship to society". That said, for the time being, growing numbers of social scientists are attracted to the idea of a distinct 'critical data studies' as marking a defiant commitment to look beyond the depoliticised boosterism and hype that propelled discussions of 'big data' and 'data science' during the 2010s. As such, critical data studies is usefully drawing together disparate critical scholarship from across the social sciences, humanities, legal and policy fields, arts and design with a declared interest in "avoid[ing] the hubris of pseudo-positivism and technological determinism, in favour of the nuanced and contingent" (Dalton *et al.* 2016, p.1).

Against this background, then, it is worth considering how critical data studies offers a set of distinct epistemological and ontological approaches from which we might make sense of future social issues relating to digital data. As previous chapters in this handbook have outlined, critical data studies foregrounds familiar social science concerns with post-positivist and interpretivist understandings, an interest in the social construction of data, and reflexive concerns with revealing and challenging dominant power structures. So, in what ways might scholarship within this tradition lend itself to anticipating possible 'data futures'? We now go on to present brief outlines of seven distinct sets of issues and concerns associated with specific bodies of literature that have emerged as part of recent critical data scholarship. Each of these, in turn, foregrounds different possibilities well worth considering in terms of how future datafictions of society might unfold.

i. Data and new forms of social order

A first set of ideas relates to the role of data in reconfiguring the overall order of social life, bringing critical sociology concerns to bear on the datafication of society and the changing nature of social order. This work raises questions about the politics of datafication and the new forms of power that are exerted through data

assemblages – not least the embodied agency of the reflexive human subject, as well as consequences of datafication for the constitution of social knowledge and the social world. As Nick Couldry (2020, p.1140) contends, approaching datafication along these lines therefore focuses attention on the crucial question of “how is the overall order of social life being reconfigured ... on the basis of new and radical forms of [data-driven] reduction?”.

Couldry expands this question by highlighting three interlinked areas of inquiry that neatly frame the imagining of any digital futures. First is the extent to which established (pre-datafied) understandings of what counts as social knowledge and who/what counts as an input to social knowledge might alter in light of ongoing datafications of society. Here, Couldry considers the example of ‘poverty’ – traditionally understood as a socially-caused phenomenon that is best addressed by giving the poor more favourable terms. In an era of big data, however, this knowledge might well be reconstituted along data-driven lines. For example, the category of being ‘poor’ might instead begin to be calculated and determined in terms of a person’s previous credit-related behaviour rather than their actual material circumstances. Moreover, the subsequent treatment of those who are designated ‘poor’ might begin to be calculated in terms of avoidance of further commercial risk on the part of organisations and institutions involved in this individual’s life. As such, whether one is designated poor or not becomes a matter of data-profiling rather than material circumstances.

Second, is how these shifts in the inputs of social knowledge might begin to change our conception of society as a whole. While automated and algorithmic knowledge might not replace existing social knowledge altogether, these new datafied models of social knowledge are likely to become more removed from older forms of expertise and judgement. In this sense, dominant conceptions of what constitutes society might begin to follow data-driven forms of social knowledge. Following on from this, is a third set of questions about how these new forms of social knowledge might begin to constitute practical understandings of what is socially actionable. This raises the need to consider how humans begin to accept the logic that social knowledge is better derived from how data-driven systems and machines interpret the world, rather than how humans might interpret things. For example, this can already be seen in the growing belief that machines are able to ‘know’ more about humans than humans know about themselves. While human expertise and judgement might not be done away with altogether, these are no longer the primary drivers of how things are done.

All told, this first set of critical data studies perspectives pushes us to think about data futures in a number of directions. This includes, the ways in which data might become an organizing principle across all aspects of society, the ways in which corporate and government actors might contrive to build different types of social order through datafication, and the values and moral dynamics of the data-driven processes that lie behind these reorganisations. In all these ways, the application of critical sociology concerns to ‘data futures’ encourages us to think about “processes of social formation themselves on the largest scale” (Couldry 2020, p.1146).

ii. Data as a form of capital

A second theme from within critical data studies relates to the changing political economy of data. Key here is the idea of data as a form of capital. This approach therefore expands on the common critical data studies contention that data is not a naturally-occurring resource. Instead, as Jathan Sadowski (2019) reasons, current forms of digital data accumulation are more accurately understood in terms of data extraction, “wherein data is taken with little regard for consent and compensation”. In this sense, digital data is taken *from* people and machines. Crucially, this process of taking data is not akin to the process of ‘mining’ a natural resource, but instead is more similar to ‘manufacturing’ an artificial object (a process which itself involves extracting vast amounts of natural resources and cheap labour).

This framing of digital data extraction is certainly evident in Thatcher and colleagues’ description of data “accumulation by dispossession” (Thatcher *et al.* 2016). This does not imply that individuals are dispossessed of their personal data in the sense of being suddenly deprived of something that they previously possessed and subsequently are not able to access (for example, in the sense of being dispossessed of one’s house or one’s dignity). Instead, rather than data being ‘taken’ from individuals, it is perhaps more accurate to talk of data being ‘transferred’ or ‘replicated’. This idea of ‘transfer’ neatly captures the role that digital technology plays in recursively transforming social actions and relations into data-based forms of property that can be appropriated in multiple (and duplicable) instances (Gray 2021).

Such distinctions raise the issue that digital data is not a commodity per se. Instead, data is more usefully seen as an asset – i.e. something that has zero marginal costs, is non-rivalrous in consumption, not profitable in its raw form but highly profitable when aggregated at scale and analysed (Komljenovic 2021). This can be seen as supporting a ‘rentier’ style of data economy, where the value of online platforms and digital devices derives from the data that they generate. An individual does not ‘own’ the devices, software and platforms that they use, and certainly does not benefit from the value of the data that is extracted from these sources of data-generation. Instead, value can only be extracted from large amounts of data that can be aggregated and analysed at scale.

Seen in these ways, then, how might we anticipate futures where data is a dominant form of capital? This idea of data-as-capital highlights various ways in which data can be used to create value: not only as a means to profile and target people, but also to optimise systems, manage and control things, model probabilities, and grow the value of assets (for example, by adding data-generating capabilities). This list evokes an idea of data futures shaped by the all-encompassing nature of this form of capitalism that is driven by a logic of continual data capture, comprehensive data coverage, and ever-expanding accumulation of data. As such, what might it mean if the actions of modern organisations are increasingly compelled by a ‘data imperative’ that demands a perpetual cycle of capital accumulation and thereby demands the extraction of all data, in every possible form, from every possible

source? As Sadowski (2019, p.8) puts it, what might we anticipate from “a world in which everything is made of data ... [and where] the supposed universality of data reframes everything as falling under the domain of data capitalism”?

iii. Data colonialism

The increased encroachment of data capitalism can be anticipated along various lines – from Shoshana Zuboff’s (2019) relatively benign and conciliatory descriptions of ‘surveillance capitalism’ through to more dystopian and oppressive visions. One useful recent line of thinking – particularly in terms of expanding our scope beyond the (over)industrialised global north – is the idea of ‘data colonialism’. Here, various authors draw equivalencies between, (i) the scale of twenty-first century capitalism being made possible through the extraction of digital data from our everyday digital practices, and (ii) previous forms of territorial colonialism that blighted the seventeenth, eighteenth and nineteenth centuries (Ricaurte 2019, Mann & Daly 2019, Couldry & Mejias 2019). While attracting criticism from some commentators concerned with obscuring the roots of current Big Tech data practices in “labour and economic structures [that] exist because of actual histories of colonization” (AI Now Institute 2019, p.44), this idea of ‘data colonialism’ frames current forms of digital datafication as a continuation of longer-term historical relationships between colonialism and capitalism. So, just as the ‘land grabs’ of the seventeenth, eighteenth and nineteenth centuries paved the way for the emergence of industrial capitalism, the current ‘data grab’ is seen as ushering in a similarly unprecedented forms of data capitalism – rooted in a new relation to the resources of the world and, it follows, a new relation to power.

In one sense, the parallels between territorial colonialism and data colonialism seem obvious, with the current political economy of digital data predicated upon ‘predatory extractive practices’, the violent imposition of a particular world-view, the exploitation of human beings as raw material, and relentless expansionism in terms of geographical space *and* the social lives of human subjects. Yet, in another sense, the idea of data colonialism implies a step-change from historical forms of capitalism - shifting the exploitation of people through labour relations into the appropriation and exploitation of their social relations. Moreover, data colonialism involves the extraction of value from many aspects of social life and society that were not previously subject to being appropriated and exploited by capitalism. Mejias and Couldry refer to this as ‘data relations’ – new types of human relations which enable the extraction of data for commodification.

This idea of data relations therefore raises a number of different futures where people are conditioned to accept (if not expect) being subjected to data-based tracking as a permanent and totalising feature of everyday life. In this sense, data colonialism is not intended as a loose metaphor, but a specific description of “a new stage of capitalism whose outlines we only glimpse: the capitalization of life without limit” (Couldry & Mejias 2019). Critical data scholars developing this notion of data colonialism therefore recognise the need to not simply theorise the possible consequences of this shift, but to also develop ways of actively resisting the forms of

data colonialism that is currently under way. In terms of anticipating possible data futures, then, this draws attention to anticipating ways in which such forms of resistance might continue and evolve over the next few decades. In short, we need to anticipate ways in which data might become a site of struggle rather than a site of surrender.

iv. Data-driven (in)justice

A fourth set of ideas relate to the challenge of making everyday engagements with digital data more equitable, empowering and 'fair'. Looking beyond talk of 'data rights' and 'data ethics' by legal scholars and philosophers, some critical data scholars have extended their interest in issues of power, disadvantage and oppression to develop the idea of 'data justice'. Such approaches mark deliberate attempts to move beyond seeing data as an individual problem and responsibility. Instead, data is framed as a collective concern that is entwined deeply with broader patterns of injustice. As Hintz and colleagues (2018, p.140) reason, this work seeks to:

“... situate datafication in the context of the interests driving such processes, and the social and economic organisation that enables them. From this, it would be possible to engage with societal implications of datafication in relation to other social justice concerns”.

These themes are expanded in Linnet Taylor's (2017) framework for data justice – reiterating the connections between data rights (the power to better engage with data) and data freedoms (the power to choose how to engage with data). As Taylor (2017, p.8) puts it, in practical terms this involves “the need to opt out of data collection or processing, the need to preserve one's autonomy with regard to data-producing technologies and the need to be protected from and to challenge data-driven discrimination”. Such rights and freedoms can be seen in terms of what Taylor identifies as 'three pillars' of data justice: (i) visibility (including access to representation, informational privacy); (ii) engagement with technology (having autonomy in technology choices, sharing in data's benefits); and (iii) non-discrimination (including the ability to challenge data bias and to prevent discrimination).

The data justice approach therefore raises a number of significant contentions about the connections between possible data futures and broader structural issues of societal inequality, discrimination and oppression. First, is the idea that data-driven social justice issues are not purely technical and/or organisational problems with likely technical and/or organisational solutions. Instead, digital data is entwined with long-standing societal problems such as discrimination, social immobility, and the social reproduction of inequality. These issues are social, cultural, political and economic – both in their origins *and* in terms of any likely responses to address them. Second, is the idea that data-driven social justice issues are not an individual concern that arise from an individual's personal choices or biases. Neither are

individuals ultimately responsible for dealing with these problems. Instead, the idea of data justice reminds us that these issues are structural. As such, any moves toward data justice require collective responses and a wide range of stakeholders being involved “in articulating both challenges and responses to datafication” (Hintz *et al.* 2018, p.136)

Third, framing datafication in terms of (in)justice prompts us to consider collective political responses and direct actions. For example, viewing data issues solely in terms of ‘rights’ can often result in responses relating to improving individuals’ consumer rights (for example, clearer and fairer terms of service), recommendations for direct consumer action (for example, individuals deleting their social media accounts), or pushes for stronger regulatory and legal protection in law. In contrast, the notion of data justice raises the prospect of direct collective actions. Data justice does not push us to imagine data futures where issues of injustice are solved by the market or by better regulation. Instead, it highlights issues that require protest, confrontation, disobedience and disruption. All told, the data justice approach evokes futures where data is explicitly recognised as a political issue, where politicisation is a necessary first step in generating the possibility of change. In this sense, the data justice approach is imbued with a spirit of collective activism and provocation. As Hintz *et al.* (2018, p.142) put it: “a wider collective moment can shake the foundations of the datafication paradigm, to reveal what else might be possible”.

v. Data feminism

This emphasis on framing perceptions of data and society in terms of structural power, oppression and inequality is also an underpinning principle of scholars working around the area of data feminism. As D’Iganizio and Klein (2020) contend:

“Feminism is about power – who has it and who doesn’t. And in a world in which data is power, and that power is wielded unequally, data feminism can help us understand how it can be challenged and changed”.

As with scholarship around data justice, data feminism explores the contention that most problems associated with current uses of data relate back to how power and privilege operate in the present moment. In many Western countries, the ways that data is extracted, utilised and benefited from are clearly bound up in white, European, male economic interests. Indeed, data feminism takes an explicit lead from Black feminist notions of intersectionality – foregrounding the idea that gender interlocks with other systems of oppression (such as race and class), prompting different people to experience different forms of benefit/harm from data regimes depending on their location within this matrix. As such, we should not set about imagining data futures where such issues might be simply avoided or neutralised. At best, we might anticipate futures where these injustices are rebalanced along more equitable lines.

Data feminism therefore explores the underpinning aims of examining and challenging the power imbalances between those who extract and those who are

extracted from – and seeking out alternate approaches that work around and counter these hegemonies. The data feminist approach therefore raises a number of pertinent issues and ideas to be taken forward into any anticipated data futures. These include principles of moving beyond essentially power-preserving ideas of ‘ethics’, ‘fairness’, and ‘accountability’, and instead foregrounding genuinely power-challenging concepts that have potential to disrupt the status quo – such as justice, ‘equity’ and ‘co-liberation’.

The idea of data justice also raises the possibility of future disruptions to data uses that are based around essentialising categories (such as false binaries and hierarchies), and instead embrace data as a means of highlighting pluralism and foregrounding conflicting viewpoints and multiple perspectives. Data feminism also promotes future scenarios where data is reconciled with emotions, embodiment and other aspects of the human experience, as well as ensuring that data is not stripped of its context – in opposition to the idea of ‘letting the numbers speak for themselves’. Finally, data feminism also highlights the need to anticipate futures where the invisible labour inherent in any data process is fully recognised and challenged. Here, specific attention is paid to low-paid and unpaid work performed by women, people of colour and outsourced non-Western workers – all of whom are integral to the functioning of the data economy.

vi. Data as a source of social good

A sixth set of ideas relates to the notion of digital data as a form of ‘social good’. Unlike the approaches described previously, the emphasis here is placed less on critique and more on optimistic explorations of possible better uses of data. This is work that is less interested in anticipating future data ‘harms’, as it is in striving for positive alternatives and hopeful visions of datafied futures. As Mann *et al* (2019, p.9) continue:

“How could data and digital technologies be designed and used in ‘good ways’, for ‘good’ purposes? If digitisation and data are inevitabilities, then we have to (re)imagine the kind of digitised world and data we want to see rather than only offering a naysaying critique of the status quo”.

There are a range of ‘good data’ practices that embody these principles. For example, the idea of data sovereignty raises the desirability of holding digital data accountable to the governance structures, cultural interests and knowledge structures relating to the communities and cultures where it originates (and to which the data relates to). In particular, this has sparked interest in issues of Indigenous data sovereignty and postcolonial forms of data – i.e. alternate forms of data processes and practices that are rooted in Indigenous belief and value systems, epistemological approaches and ontological assumptions. As Walter and Suina (2019) note, this foregrounds the need for different forms of data assemblage that produce data that is disaggregated, that disrupts deficit narratives, that advances

Indigenous nations' re-building agendas, and is rooted in arrangements where access to this data is controlled by Indigenous representatives.

Another key line of 'good data' thinking is the development of democratic forms of data stewardship and data sharing approaches to creating 'value' from data – shifting focus away from economic revenue to values of public good, social solidarity and citizens' self-determination (Micheli *et al.* 2020). In practical terms, this suggests efforts to enact these principles through the establishment of 'data cooperatives' by civic organisations – where individuals' data is aggregated in the form of data trusts, data sharing pools and other forms of data commons that might then be used for socially-just purposes that are collectively decided upon. At a municipal level, these principles also anticipate public bodies establishing 'public data trusts' – collating citizen's data to better inform policy-making and promote the idea of 'data as a public infrastructure'

The good data approach also encompasses a range of practices that is sometimes referred to as 'data activism'. This involves groups that are traditionally marginalised and disadvantaged by dominant data regimes working to make alternate use of digital data for their own self-defined political ends (Milan and van der Velden 2016). Such activism might involve the repurposing of existing data and/or the generation of new data – all with the intention of contending and disputing official accounts, drawing attention to underpublicized issues, and mobilising public opinion. Data activism can involve the preservation of existing data-sets, and collective efforts to actively generate new forms of data. Such uses of data and statistics as a resistant tool has been described as 'statactivism' (Bruno *et al* 2014) – a neat term that encompasses the reuse of data about people and places by the communities that these data purport to represent. In all these ways, then, good data raises visions of data futures that are hopeful and collectively empowering.

vii. Data as a site of resistance

Finally, is a set of ideas that expands the need for radically confrontational and resistant approaches to countering the harms that data can do to traditionally marginalised, disadvantaged and oppressed groups. This includes emerging literatures around 'queer data', 'black data' and 'crip technoscience' – all of which strive to develop more radically-minded approaches toward the countering harms that arise from dominant data regimes. Scholars working in these areas make deliberate use of the term 'minoritized' to describe the groups most disadvantaged and oppressed by dominant uses of data – thus foregrounding unequal power relations that, in many parts of the world, have historically seen the dominance of white European male interests in determining and maintaining the public sphere. One starting point for these approaches is acknowledging the double-bind of how digital data renders minoritized groups 'hypervisible' to authoritarian surveillance, while at the same time ensuring that minoritized people do not 'conform' with dominant data regimes. Thus digital data serves to both misrecognise and oppress these groups in ways that make conventional forms of digital data profoundly inappropriate and therefore something to be wholly rejected and resisted.

This sense of data rejection and resistance is evident in the idea of 'black data' (McGlotten 2016) – critical scholarship focused explicitly on the inherently racialized dimensions of data culture, economics and politics - challenging the ways in which 'whiteness' is continually affirmed as normative across the most dominant features of datafication. The idea of black data starts from the premise that any attempt to datify people along lines of race is deeply problematic for a number of reasons – not least the fact that 'race' is not a hard-and-fast 'variable' relating to biological traits, but a socially constructed notion. In this sense, black data scholars remind us that digital racial profiling directly links back to historical forms of racialized profiling that are long discredited and acknowledged to be discriminatory. Key themes within the black data approach include how data-driven systems are used to misrepresent and segregate people along lines of race, all leading to outcomes that discriminate and oppress (Gandy 1993, 2019).

This fundamental incompatibility between the prevalence of digital data and the everyday experiences of black, queer, crip and other minoritized groups therefore leads to the provocative move of anticipating futures where conventional assumptions around digital data and data science are rejected out of hand. Instead, expectations and norms that currently surround data and datafication are to be challenged and resisted – particularly in terms of challenging how digital data is entangled with power relations and the politics of oppression (Jakobsen 1998). Key here is the understanding that none of these dominant conditions of data use can be easily altered for the better. Instead, these ideas foreground the idea of radical rejection and resistance toward dominant data practices and products. This is reflected in the growing academic (and activist) interest in 'data obfuscation' and 'vernacular resistance' to data collection and analysis – i.e. tactics of 'data disobedience' intended to mitigate, evade and sometimes sabotage dominant structures of data reuse and recirculation (Brunton & Nissenbaum 2015). All told, the act of actively disrupting the production and processing of data is seen as a powerful means of increasing minoritized groups' agency and autonomy, as well as reducing the visibility of these groups - what McGlotten (2016, p.269) describes as "black-boxing ourselves to make ourselves illegible to the surveillance state and big data".

At the same time are attempts to reimagine data along more confrontational lines. Writing from a critical race perspective, Crooks and Currie (2021) advance the idea of 'agnostic data' – i.e. taking a conflict (rather than consensus) approach toward data activism where digital data is reappropriated in adversarial ways. This involves using data to highlight difference, raising (rather than resolving) questions, and stimulating grassroots political involvement and engagement with alternative futures. The idea of data not being used to provide answers, but instead being used to provoke questions and strengthen solidarity therefore raises some radically different forms of possible data futures.

2. THINKING ABOUT 'DATA FUTURES': FROM CORPORATE HYPE TO ACADEMIC ANTICIPATION

These various 'lenses' from recent critical scholarship around data and datafication raise various logics, ideas and concerns that might play out in future forms of data-driven society. Next, then, we need to align these ideas emerging from 'critical data studies' with the recent turn within the digital social sciences toward 'futures thinking', anticipation studies, sociotechnical imaginaries, the sociology of expectations, and other forward-facing perspectives of the digital-society-to-come.

As the concluding section of this handbook implies, digital social science is becoming steadily in thrall to the idea that futures – rather than histories – are the most useful means of making sense of the digital present. This picks up on the increased interest that social scientists have shown in thinking about futures since the 1980s. What was once seen as a rather spurious area of scholarship (e.g. 'futurology') has gained considerable credence over the past forty years or so as an essential element of contemporary social theorising and thinking. Building on initial work from the likes of Wendell Bell and Wilbert Moore in the 1970s, the past twenty years has seen serious thought being given to future-looking forms of social science – from Barbara Adam's (2007) work on futures thinking through to scholarship around the narrative and discursive construction of techno-science futures (Brown and Rappert 2000). As the sociologist John Urry (2016, p.187) challenged his fellow social scientists, "in order to operate in the world ... we must peer into the future – there is no choice".

Notwithstanding this general turn within the academic social sciences, most talk of future forms of datafication remains dominated by commentators content to anticipate data futures in decidedly non-critical terms of linear progress. For many people, such extrapolations make good common-sense. After all, the technological classes consider themselves to be keenly attuned to what areas of current 'emerging' technology development are likely to thrive over the next few decades. For example, it seems reasonable to expect that the next few decades will see continued innovation in areas such as artificial intelligence, automated decision-making, big data analytics, 'edge' computing, platformed personalization, smart sensors and the 'internet of things'. It might also be the case that continued work in neuroscience and pharmaceutical technology leads to significant break-throughs in cognitive analytics and emotion detection, along with the development of DNA-based data storage allowing the storage of huge volumes of digital data at a molecular level.

Yet, even if these areas of technology development do continue as broadly expected, this tells us little about the societal impacts that such advancements might have. As such, it makes sense to be wary of anyone who is too confident in their claims about the ongoing datafications of society. Indeed, the most bold and assured visions of digitally-driven futures tend to be made by people with something to sell. A real-estate developer planning a futuristic city-state in the Saudi desert has a vested interest in pushing visions of a data-driven 'smart city' with automated flying-taxis and constant facial recognition monitoring where "everything can be recorded"

(Scheck et al. 2019). More prosaically, the managing director of a major bank's digital innovation unit is understandably eager to foresee a future where banks become trusted brokers of people's data to third parties and take responsibility for 'streamlining' issues of consent (Norton-Smith 2021).

Of course, not all talk of future data trends are quite so self-interested, yet even seemingly more objective attempts at mapping empirical trends tend to present hollowed-out and highly speculative accounts. Preferred mechanisms for such forms of future-telling tend to be predictions, forecasting and other methods of 'accurate' speculation. Here, for example, we find numerous corporate proclamations of impending expansions of the data economy and data market – involving breathless talk of hundreds of 'zettabytes' of data being produced each year, and exponential expectations such as "the amount of data created over the next three years will be more than the data created over the past 30 years" (Press 2020, n.p).

Again, such visions are presented as rooted in plausible extensions of likely linear technological progress. For example, most experts concur that pronounced increases in data production will occur as a result of the rise of embedded data from sensors – especially harnessing data from live video and audio streams. For example, Gartner (2021) estimate 75 percent of workplace conversations being recorded and analysed by 2025. This growth is expected to be fuelled by the datafication of low-income countries – especially expansions into regions with previously low levels of internet connectivity. Other expected trends include the attendant rise of data-related 'cyber-crime' (i.e. crime related to data as a valuable asset), as well as the increased use of 'synthetic' data (i.e. data modelled on 'real-life' data) to provide sufficient volumes of data to train deep learning AI models. All told, there are various rationales from the world of technology and business to presume that our data futures are already set on a fairly predictable path.

Yet, while not as spurious as envisioning 'smart cities' deep in the Saudi desert, such predictions remain speculative and lacking in social context. To be blunt, these are technology and business-led visions that lack social curiosity, compassion, or concern. Of course, there is merit in digital social scientists engaging with such predictions as a way of examining the discursive construction of techno-social conditions by corporate actors, and how such predictions act as a means of providing commercial actors with legitimacy. Indeed, from a 'sociology of expectations' perspective, engaging with corporate and policy predictions can be a 'fundamentally 'generative' way for social scientists to examine how corporate actors use ideas of the future to "guide activities, provide structure and legitimation, attract interest and foster investment. They give definition to roles, clarify duties" (Borup *et al.* 2006, pp.285-286). As such, there is a growing critical literature addressing the 'sociotechnical imaginaries' being advanced by governments, corporations and other stakeholders in the data society and data economy (e.g. Sadowski & Bendor 2019, Means 2018).

Nevertheless, we need to remain mindful of recent criticism of these humanities and social science researchers tempted to play along too closely with the dominant forecasts and predictions that inform mainstream understandings and conversations

about the digital/data future. Recently, Lee Vinsel (2021) has derided the “academic business model” of critical researchers “playing along with hype to score cash money and prestige” - what Vinsel curtly terms “lend[ing] credibility to industry bullshit”. This chimes with Alfred Nordmann’s (2007) concerns with social scientists who develop critiques of ‘if/then’ scenarios that engage credulously with dramatic corporate claims over ‘incredible futures’ driven by ‘technological hubris’. The dangers of engaging with speculative digital futures work of this type are obvious. Engaging critically with corporate expectations of data futures risks reproducing (and even increasing) hyperbole, and lending credibility to industry promotional claims simply by taking them seriously, overstating the abilities of tech firms and/or the capabilities of their emerging products. As well as compounding unrealistic expectations, Nordmann (2007) warns that work of this nature distracts attention away from current already-existing concerns - especially the actual risks and harms arising from current existing technologies. Given the thrust of the various critical data studies ‘lenses’ outlined previously, engaging with the topic of ‘data futures’ without giving due attention to matters of injustice, inequality, and resistance makes little sense.

So, how else might the social sciences engage more usefully with the question of ‘data futures’? As has just been argued, current dominant future visions not only suffer from a lack of social/societal awareness, but also fail to approach the datafication of society in sufficiently problematic terms. More specifically, these current data futures are limited by underlying expectation of a continuation of current conditions, and the presumption that these conditions are largely beneficial. Regardless of their thrust, all these forecasts and predictions tend to presume that fundamental elements of current data society, data economy and data politics will continue. In short, it is presumed that some form of the status quo will endure, even if we cannot be sure exactly in what form.

Approaching the future of data and society through this lens of what seems ‘probable’, ‘likely’, and ‘familiar’ might seem sensible in comparison to wildly speculating about killer robots and colonies-on-the-moon. Yet, from a social science perspective, approaching data futures in terms of the probable, likely and familiar remain decidedly limited starting-points - foreclosing alternative ways of doing things that might prove necessary in light of unforeseeable conditions and circumstances. Instead, one of the key aspects of social science approaches to futures thinking is the incorporation of what is unforeseeable alongside what is foreseeable. This attention to the unforeseeable underpins the tendency for social scientists to talk about ‘futures’ in a plural sense. This emphasis on ‘futures’ (plural) acknowledges that there is no one linear future that can be predicted and forecasted. Instead, it makes better to prepare for possible different futures which may (or may not) occur. The emphasis here is on preparedness rather than prediction – acknowledging that many future risks such are literally unknowable (Beck 2009). As John Urry (2016, p.1) puts it: “futures are unpredictable, uncertain and often unknowable”.

This spirit of hopeful (re)imagining in the face of unknown futures is reiterated in the recent turn toward ‘anticipatory’ approaches in futures studies (see Poli 2017, Miller 2018). The contention here is that unknowable futures demand responses that imagine how we might like to be living, particularly in terms of re-centring the

perspectives of previously marginalized interests and non-powerful groups (Facer 2019). In light of the critical data ‘lenses’ outlined previously, this therefore raises a host of challenging questions that could be asked of datafication in the 2020s. For example, how might alternate approaches to living with data be established that do not presume the continuation of dominant ‘Big Tech’ industry, capitalist modes of production, the ‘global middle class’, or other facets of neoliberal hegemony? Crucially the focus here is on using these imaginations to inform ‘anticipatory behaviours’ – i.e. changes in our present behaviours rooted in these imagined and desired futures (Poli 2017).

3 - ANTICIPATING FORMS OF DATA AND SOCIETY FIT FOR UNKNOWABLE FUTURES

So, how might we anticipate desirable forms of data and society in light of essentially unknown futures? In this spirit, we can outline the brief beginnings of two scenarios that illustrate the scope of what different critical data futures might look like. Scenario-building is one of a range of ‘foresight’ (as distinct from ‘forecast’) approaches that lend themselves to developing understandings of possible futures of data and society. Here, the construction of scenarios aims to construct vignettes of different ways that our digital societies might develop, and then identify key factors that would be capable of driving change in these different particular directions.

While concerned with broadly the same types of data technologies, these different scenarios highlight the contested nature of what might be considered as plausible data futures. Clearly, how these scenarios might be elaborated depends on whatever different forms of society, shared values and political philosophies we see as preferable (e.g. social democracy as opposed to libertarianism, communitarianism as opposed to capitalism). Various iterations of these basic scenarios might be developed that reflect different relationships between government and commercial actors, and different cultural understandings about community, family, work, and other institutions. These are not accurate predictions of what is likely to happen. Instead, these are expressions of what we might like to happen:

Scenario#1. Resisting a rampant data economy

This first scenario anticipates the continued expansion of the data capitalist logics described earlier – i.e. the acceleration and intensification of dynamics of data extraction and exploitation. So, what might societies predicated around the ‘datafication of everything’ look like? Here, we might imagine the rearrangement of public services around the logics of platform rentierism – for example, health-care provision where patients do not visit a local doctor but are triaged through the proprietary platform provider that happens to be in partnership with their local municipality. This data is then shared with public health agencies as well as being sold onto various commercial third parties. Elsewhere, we might also imagine the addition of data-generating capabilities to public assets in order to grow their economic value – such as ‘smart libraries’ or ‘datafied parks’. Of course, some public

spaces and services might simply prove unprofitable to extract data from, and therefore cease to be maintained. Hard-to-quantify aspects of everyday life might be usurped by crude 'proxies' and profiles - reduced to a set of more easily generated data-points that stand-in for their more complex precedents.

This scenario also anticipates how people might experience these data-driven intensifications. This might include feeling compelled to adjust one's patterns of living and working in ways that 'fit' with these data regimes – in other words people having to behave in ways that are easily datafiable and 'algorithmically visible'. Such adjustments might involve speaking in easily parsed dialects and tones, contorting one's face into decodable expressions, and generally behaving in standardised, extractable ways. It is also important to consider what forms of social knowledge will be privileged and valued under these conditions. For example, people's social status might be grounded in predicted behaviours and probabilistic risks, rather than past accomplishments and actual achievements. On the flipside, new forms of 'coded inequity' might emerge in the form of data-driven racism, classism, ableism, and so on.

Of course, such data futures might understandably appear decidedly undesirable to many social scientists. As such, it is worth considering how these conditions might be negotiated and resisted by those *not* directly profiting or benefiting from the extraction of data. This could take the form of mass 'data obfuscation' and 'data disobedience' – with people disrupting dominant regimes of data collection and analysis in ways that damage data capitalism. Such actions might range from minor acts of infrastructural interference through to mass acts of data refusal and data contamination. We might also agitate for organised resistance on the part of tech workers and other sympathetic data industry 'insiders'. Tech-savvy citizens might work on the development and open distribution of counter-datafication tools – such as encryption tools, geo-spoofing software, VPNs and other 'off grid' technologies. Most provocatively, we might see minoritized groups establishing alliances with the interests and orientations of other resistant groups such as anarchists and cryptographers in order to develop creative forms of 'black ops' (McGlotten 2016)

At the same time, we might anticipate future forms of adversarial reapplications of data – collective uses of data that promote dis-satisfactions with dominant datafication regimes, that provoke citizen resistance, and inspire people to engage with alternate forms of data use. This might see digital data being reappropriated to publicise the differences, contradictions and absurdities of the data-driven organisation of society - provoking people's passions and directing public anger against datafied society and data capitalism. In the words of Crooks and Currie (2021, n.p), digital data might provide a means of affective stimulation to "enflame political differences and mobilise communities to begin to act" against the data economy.

Scenario#2. Rethinking datafication in an era of climate crisis

This second scenario considers possible data futures in light of the multiple, interconnected crises associated with ongoing environmental instabilities. This encompasses a range of issues: from the part that data extraction plays in the depletion of non-renewable natural resources through to the deleterious human and ecological consequences of anthropogenic climate change. These constitute pressing areas of uncertainty over the next few decades, and clearly need to feature in any talk of the futures of data and society. So what possible data futures might unfold in an era of climate catastrophe and ecological instability? What might it mean to take seriously “data’s entanglements with fossil fuels [and] the oft-neglected ecological impacts of computational capitalism” (Taffel 2021, p.9)?

These questions anticipate a state of digital ‘post-abundance’, with a planet that can no longer sustain the energy demands of unlimited ‘cloud’ storage and data processing, the energy demands and natural resource depletion implicit in the production of digital hardware as well as the damaging effects of ‘e-waste’ and digital disposal. In short, these are data futures predicated around the challenge of what to do once the world passes the point of ‘peak tech’ – facing up to the “need to abandon the fantasy that data extraction is a weightless endeavour that can grow infinitely” (Taffel 2021, p.14). In blunt terms, then, how might we anticipate data futures that are not built upon continuous and plentiful extraction of data?

Here, it is possible to imagine scenarios where data capitalism is forced to reconsider how economic value might be most efficiently generated by an ever-dwindling number of digital assemblages. This might involve tech firms concentrating on only the most profitable forms of data extraction and processing – with profitable individuals and organisations recompensed for the continued datafication of their everyday activities, while the majority of others are cast aside and rendered unquantifiable. This might see a reordering of society around a hierarchy of ‘data classes’ based around the economic value that can be generated from their datafication. At the same time, only the most potentially profitable data gets archived and stored. Digital data itself becomes a rare precious resource.

Alternately, we might imagine an era of ‘digital de-growth’ based around logics of data justice and good data – where the value of data is framed in terms of social benefit and public good. This might necessitate the immediate cessation of resource-greedy data practices that lead only to economic profit rather than social good. For example, there would be no sense in supporting continued cryptocurrency mining, or continuing to extract data simply to target advertising toward consumers to purchase non-essential products. Instead, we might choose to redistribute any still-available digital and data resources in ways that deliberately advantage the global poor and those displaced by climate migration. We might prioritise what data processing capacity remains for climate and ecological modelling, and other data applications that might help communities cope with ongoing environmental degradation. We might choose to prioritise data for purposes of public health. Alternately, we might imagine autonomous local collectives working outside of state and market systems to generate their own data that relates directly to the immediate effects of flooding, drought, winds and other unpredictable and highly localised conditions. As Sy Taffel (2021, p.12) speculates:

“This growth should be achieved alongside substantial reductions in current digital activities that serve little or no social or ecological purpose. What constitutes social benefit will be highly contentious and requires significant debate including input from marginalised groups; however, one productive approach involves re-evaluating practices based upon use values, rather than the currently hegemonic practice of leaving this to market-based exchange value”.

CONCLUDING DISCUSSIONS

How we chose to anticipate data futures depends upon how unknowable we acknowledge these futures to be. Clearly, in this current era of global economic and political instability, ongoing climate change and potential future pandemics, it seems sensible to adopt a starting-point that the future is never wholly predicable and not set simply to unfold in broadly similar ways to what has come before. Clearly, any data futures that we wish to anticipate from a social science viewpoint should not be seen as straightforward extrapolations of present predicaments. Instead, we should remain open to the idea that datafication might unfold in ways that are essentially uncertain, multi-factorial and non-linear in nature. The futures of data and society are embedded within various complex predicaments and challenges – a fact that should prompt us to think about “futures [as] made up of unstable, complex and interdependent adaptive systems” (Urry 2016, p.188).

Of course, the main goal of addressing any possible data future is to make better sense of how we might want to act in the present. Thus, the two scenarios just presented are perhaps most useful as a source of ideas for interpreting current unfolding data-related social and technical change. As Keri Facer (2020) puts it, ideas of ‘the future’ are primarily ideas to be used in the present – prompts to see new possibilities, ask different questions, and establish new practices. This thought is echoed in Richard Barbrook’s (2007, p.8) contention that “the future illuminates the potential of the present”. Thinking about future forms of data and society can therefore prompt us to move on from old ideas and assumptions. As such, we need to “think about how different ideas of the future create new possibilities and resources to stimulate, challenge and inform what we are doing in the present” (Facer 2020, n.p).

In this sense, one of the primary benefits of continuing to engage further with the idea of critical data futures is the ways in which we are pushed into re-imagining what data and datafication might be under different circumstances and conditions. Most usefully, the idea of critical digital futures pushes us to consider what present actions are required to order to fulfil these visions. Indeed, as Nordmann (2007, p.42-43) puts it, we should take “such scenarios seriously enough to generate insights from them and to discover values” that help us reflect upon current societal arrangements. At the same time, however, Nordmann reminds us to not “take them seriously enough to believe them”. Engaging with the two futures scenario just presented (and the critical data studies ‘lenses’ that they extend) therefore raises values of what might

constitute 'fair' use of data, approaching the harms of data along relational lines, and so on. Engaging with these futures scenarios also raises logics of co-operativism, de-growth and political resistance. Engaging with these futures scenarios draws attention to reconfigurations of regulation, governance, resistances and responsibilities that perhaps need to be foregrounded in terms of current debates around the ongoing datafication and data-led shaping of society. Above all, these scenarios bring a welcome focus to bear on people (in particular, people's interests, relations, and diverse life circumstances and everyday experiences), reminding us that any sense of 'data futures' needs to be framed as human futures. As Genevieve Bell (2021, n.p) contends:

“In this moment, we need to be reminded that stories of the future ... are never just about technology; they are about people and they are about the places those people find themselves, the places they might call home and the systems that bind them all together”.

In all these ways, then, continued discussion of possible data futures offers a neat way for critical data studies to keep an eye on 'how might things be otherwise'. Of course, there are limits to this endeavour. For example, we should not lose sight of the lessons that can be learnt from data histories. Indeed, many of the data future scenarios just outlined are rooted in pre-digital histories and continuities – not least ideas of co-operativism, activism, community organisation and citizen resistance. Just as critical data studies are quick to criticise 'Big Data' advocates as presenting their field as “perpetually new, ahistorical and revolutionary” (Dalton *et al.* 2016, p.3), neither should the digital social sciences fall into a similar trap when outlining alternate possibilities. Moreover, we should not lapse into lazily imagining overly-generic scenarios, and lose sight of the importance of local context, local conditions and how data are shaped by the places they are produced in.

These warnings notwithstanding, the idea of critical data futures provides a ready means to put the idea of the data-driven society under pressure. In this spirit, the approaches outlined in this chapter leave digital social scientists with plenty of work to. On one hand we need to continue to develop what Ruppert (2018) terms “counter-imaginaries of datafication” – fighting currently prevalent corporate ideas of what is thinkable about data and society – not least dominant understandings of what practices are preferable and which actors we might want to be performing them. At the same time, we need to work to convert these preferable futures into a workable present. As Kazansky and Milan (2021) suggest, this includes moving our thinking from currently speculative responses through to more formative plans, and then on to stabilised sets of tools and tactics. We also need to pay full attention to any accompanying ontological and epistemological changes that might be necessary in our social institutions and society at large. Above all, we need to retain hope that the dominant forms of datafication currently to be found in our digital societies *not* a done deal. Instead, we need to remain inspired by our collective capacities to anticipate better data futures to come.

REFERENCES

- AI Now Institute (2019) *2019 Report*. AI Now
https://ainowinstitute.org/AI_Now_2019_Report.pdf
- Barbrook, R. (2007). *Imaginary futures*. Pluto Press
- Bell, G. (2021). Touching the future. *Griffith Review*, 71,
www.griffithreview.com/articles/touching-the-future/
- Borup, M., Brown, N., Konrad, K. and Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, 18(3-4), 285-298.
- Brown, N. and Rappert, B. (2000). *Contested futures*. Routledge.
- Bruno, I., Didier, E. and Vitale, T. (2014). Statactivism. *Partecipazione e conflitto*, 7(2), 198-220.
- Brunton, F. and Nissenbaum, H. (2015). *Obfuscation*. MIT Press.
- Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, 22(7), 1135-1151.
- Couldry, N. and Mejias, U. (2019). Data colonialism. *Television & New Media*, 20(4), 336-349.
- Currie, M. and Paris, B. (2018): Back-ups for the future. *Archives and Manuscripts*, 46(2): 124-142.
- D'Ignazio, C. and Klein, L. F. (2020). *Data feminism*. MIT Press.
- Dalton, C., Taylor, L. and Thatcher, J. (2016). Critical data studies. *Big Data & Society*, 3(1), 2053951716648346.
- Facer, K. (2019). The university as engine for anticipation. in Poli, R. (ed) *Handbook of anticipation*. Springer (pp.1439-1457)
- Facer, K. (2020) *Eleven things I've learned from educational futures work that might be helpful for COVID-19*. March 31st,
<https://kerifacer.wordpress.com/2020/03/31/11-things-ive-learned-from-educational-futures-work-that-might-be-helpful-for-covid-19/>
- Press, G. (2020) 54 predictions about the state of data in 2021. *Forbes*, 30th December
- Gandy, O. (1993). *The panoptic sort*. Oxford University Press.
- Gandy, O. (2019). Data mining, surveillance, and discrimination in the post-9/11 environment. in Haggerty, K. & Ericson, R. (Eds) *The new politics of surveillance and visibility* (pp. 363-384). University of Toronto Press.
- Gartner (2021). Top ten strategic predictions for 2021.
www.gartner.com/smarterwithgartner/gartner-top-10-strategic-predictions-for-2021-and-beyond/
- Gray, C. (2021). *Data dispossession*. <https://www.thesociologicalreview.com/data-dispossession/>
- Hintz, A., Dencik, L. and Wahl-Jorgensen, K. (2018). *Digital citizenship in a datafied society*. Polity.
- Jakobsen, J. (1998). Queer is? Queer does? Normativity and the problem of resistance. *GLQ: a Journal of Lesbian and Gay Studies*, 4(4): 511-536.
- Kazansky, B. and Milan, S. (2021). Bodies not templates. *New Media & Society*, 23(2), 363-381.
- Komljenovic, J. (2021). The rise of education rentiers. *Learning, Media and Technology* [forthcoming]

- Mann, M. and Daly, A. (2019). (Big) data and the North-in-South. *Television and New Media*, 20(4), 379–395.
- Mann, M., Devitt, S. and Daly, A. (2019). What is (in) Good Data? in *Good Data*. Institute of Network Cultures Theory
- McGlotten S (2016) Black data. in: Johnson P. (ed.) *No tea, no shade: new writings in black queer studies*. Durham: Duke University Press, pp.262–286
- Means, A. (2018). Platform learning and on-demand labour. *Learning, Media and Technology*, 43(3), 326-338
- Micheli, M., Ponti, M., Craglia, M. and Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2), 2053951720948087.
- Milan, S. and Van der Velden, L. (2016). The alternative epistemologies of data activism. *Digital Culture & Society*, 2(2):57-74.
- Miller, R. (2018). *Transforming the future: anticipation in the 21st century*. Routledge
- Noble, S (2018). *Algorithms of oppression: how search engines reinforce racism*. NYU Press.
- Nordmann, A. (2007). If and then: a critique of speculative nanoethics. *Nanoethics*, 1(1):31-46
- Norton-Smith, T. (2021). Can banks be custodians and brokers in the data economy? in *30 Voices on 2030 - The New Reality of Financial Services*. KPMG (pp.76-77) <https://assets.kpmg/content/dam/kpmg/au/pdf/2021/30-voices-on-2030-new-reality-financial-services.pdf>
- Poli, R. (2017). *Introduction to anticipation studies*. Springer.
- Ricaurte, P. (2019). Data epistemologies, the coloniality of power, and resistance. *Television and New Media*, 20(4), 350–365
- Ruppert, E. (2018). *Sociotechnical imaginaries of different data futures*. Erasmus School of Social and Behavioural Sciences
- Sadowski, J. (2019). When data is capital. *Big Data & Society*, 6(1), 2053951718820549.
- Sadowski, J. and Bendor, R. (2019). Selling smartness. *Science, Technology, & Human Values*, 44(3), 540-563.
- Scheck, J., Jones, R. and Said, S. (2019). A Prince's \$500 billion desert dream. *Wall Street Journal*, July 25th
- Stevens, N. and Keyes, O. (2021). Seeing infrastructure. *Cultural Studies*, [forthcoming]
- Taffel, S. (2021). Data and oil. *New Media & Society*, 14614448211017887.
- Taylor, L. (2017). What is data justice? *Big Data & Society*, 4(2), 2053951717736335
- Thatcher, J., O'Sullivan, D. and Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession. *Environment and Planning D: Society and Space*: 34(6): 990-1006.
- Urry, J. (2016). *What is the future?* Cambridge, Polity
- Vinsel, L. (2021). You're doing it wrong: notes on criticism and technology hype. *STS News*, 2nd February, <https://sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5>.
- Walter, M. and Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22(3), 233-243.