Introduction: Towards an Anthropology of Data

Rachel Douglas-Jones, Antonia Walford and Nick Seaver

Abstract

The world is talking 'data'. The early cross-disciplinary, business-oriented hype around the potential of 'big' data, with its promises of unprecedented insight into social life, has given way. Data now motivates a sweep of dystopian visions, from rampant commodification, to the invasion of privacy, political manipulation, and shadowy data doubles. Yet anthropologists have been cautious in taking data itself as their object, even as the social life of data practices becomes manifest in our ethnographies. In this introduction, we argue for an anthropology of data that is ethnographically specific and theoretically ambitious, putting forward a case for why anthropological engagements with the data moment might be not only politically important but also conceptually generative.

Among experts and amateurs, from the spectacular to the mundane, something called 'data' has permeated the ethnographic field. In Yogyakarta, Indonesia, a taxi driver tells an anthropologist how he puzzles over the booking apps that allocate rides to him; he considers putting his accounts into 'therapy', strategically accepting and declining requests so that he will have 'better data' (Sandbukt 2020). On the border between Kyrgyzstan and Russia, an ethnographer studying seasonal migrant labor finds herself concerned by data gathered into the 'black list', a tool used by Russian immigration services to make migrants deportable (Reeves 2016). In the United States, a grieving daughter describes how a personal database of memories appears to bring her mother back to life (Hales 2019). From conflicting data about voter turnout in the 2019 Indian elections, apparitions of 'ghost voters' emerge (India Today 2019). As a viral pandemic encircles the globe,

data about infection and mortality rates becomes a matter of international public dispute (Street and Kelly 2020). Anthropologists find themselves needing to make sense of ‘data’ - and not only their own - as an emergent ethnographic object.

At a moment when ‘data’ commonly implies a universalizing epistemology, the papers in this volume attend to data’s multiplicity and particularity, relocating it in diverse local worlds. The growing presence of data in our field sites demands this traditional ethnographic work, which foregrounds the practical existence of data, in small forms that complicate and exceed its ‘big’ reputation. This collection starts from the popular association of ‘data’ with digital technologies – networked computing systems that register, store, and analyze ever more information about ever more aspects of life – but it also expands the frame, looking to data’s relations with existing informational forms, such as documentation and accounting, to find continuity, rather than disjuncture. The work collected here charts a course between hailing data as a radical rupture with the past, and recognising data as simply a continuation of familiar practices of social ordering, by attending to how discourses, practices and imaginaries of data are configuring and inflecting the familiar in unfamiliar or surprising ways.

In this special issue, we draw data’s apparent novelty into conversation with many of anthropology’s central concepts, from kinship to value to personhood. This Introduction demonstrates the necessity of this theoretical project, describing one shape it could take. Our first section is diagnostic, disambiguating different ways that data is talked of and done. We illustrate data’s charismatic hold and review key approaches to its study within anthropology so far. Looking to an earlier moment of technological and conceptual innovation – the advent of new reproductive technologies and the anthropological response to them – we identify areas of analytical interest we might attend to in the present data moment.By treating data as at once an empirical concern for ethnographers and also as an opportunity to revisit key anthropological concepts, the pieces collected in this volume show the potential for a transformative anthropology of data – one that goes beyond updates to the ethnographic record and uses data as a generative site of anthropological theory-building.

Diagnosis and Disambiguation in the Data Moment

Data, despite its apparent simplicity in common use, is not simply 'given' (as the word's etymological roots might suggest). What it is, and what it means, shifts. Everyday definitions of data start with the familiar: documents, numbers, building blocks of quantification, governance and analysis. However, borrowing charisma from computation, data has also acquired a more revolutionary reputation: a powerful entity that shapes commercial futures, an engine of growth that drives institutions to 'datafy' themselves in pursuit of profit or efficacy (Fourcade and Healy 2017). To make sense of data, we need to first disambiguate the many uses of the term.

Knowing Data

One reaction to data's ubiquity and polysemy is to insist on data's material specificity. 'Data' continues to designate concrete, straightforward, empirical stuff (Dourish 2017). It takes the form of spreadsheets, .csv files, digitized archives, hand drawings and notes, or graphs on laptops; it is collected through sensors, database entries, mobile phones, and census work, from questionnaires to clicks. Thus understood, we can trace a decades-long history of ethnographic work on data, including the nature of evidence (Hastrup 2004; Engelke 2008), governmental practices of quantification and accounting (Maurer 1997; Nelson 2015), and the production of indicators by international non-governmental organizations and experts (Merry 2016; Murphy 2017).Data has appeared in the ethnographic record in analyses of 'audit culture' in the 1990s (Power 1997; Strathern 2000), studies of the charismatic role of documents in bureaucracy (Riles 2006; Hull 2012), and the seductive power of the numerical (Porter 1995; Verran 2010; Merry 2016).

Anthropologists and historians have also drawn out the importance of colonial bureaucratic practices of enumeration, which underpinned territorial dispossession and enforced a violent legibility on colonised peoples (Appadurai 1994; Scott 1998); more broadly, European empire-building relied foundationally on quantification and the 'avalanche of numbers' that captured the 18th and 19th C European imagination, and the standardisation, classification and accounting that it permitted (Poovey 1998; Hacking 1990).[i] These continuities suggest that data is in fact a familiar

concern for anthropologists, knowable through already-existing frames. The data revolution may not be so revolutionary after all (cf. Pfaffenberger 1988).

However, data has come to signify and act beyond these earlier practices. As people around the world reckon with data's significance, we re-open the definitional question as an ethnographic problem: What is data? As an object of ethnographic scrutiny, data is not merely varied, it is mercurial. What it is can change depending on the use to which it is put, or when one asks (Star and Ruhleder 1996). Data might be a store or source of value, an asset or a liability. Data for the taxi driver in Yogyakarta is not same thing – either as an object or as meaning – as data for a worker placed on a Russian 'black list'. Tom Boellstorff and Bill Maurer describe this mercurial character as a consequence of data's sociality:

> data is formed through relations that extend beyond 'data' itself; [...] what counts as data (and data's referent) is a social process with political overtones; and [...] data is always in real-time transformation in ways that cut across notions of nature and culture. (Boellstorff and Maurer 2015: 3–4)

These relations, processes, and transformations are at the heart of this special issue. Understanding data as socially constituted has consequences for how we understand both academic disciplinary and public responses to data's growing prominence.

Data and Disciplines

For many, 'data' appears to transcend old disciplinary boundaries and to promise new synthetic knowledges[ii]. For others, this universalist drive represents a threat to the particularity of disciplinary expertise. Sociologists have described the growth of 'social' big data collection by commercial firms as a threat to their empirical authority (Savage and Burrows 2007); geographers, worry about being 'left behind as others leverage insights from the growing data deluge' (Kitchin

2013).Social scientists have pursued data as a means to know the social (e.g. Tinati et al. 2014; Marres 2017). Historians have set about applying digital methods such as text mining, visualization and mapping techniques, to their topics (Mullaney 2019a, 2019b), and humanists continue to pursue digital, data-driven projects (Terras, Nyhan and Vanhoutte 2013; Gardiner and Musto 2015). These responses continue a history of promise and threat that dates back at least to the advent of computing itself (see Seaver, this volume).

What, then, of anthropologists? With some exceptions (e.g. Crowder et al. 2019; Madsen et al 2018; Knox and Nafus 2018), anthropology has largely responded to the data moment by figuring ethnographic fieldwork as a necessary or more sensitive qualitative complement to large-scale data collection and analysis. It has done so in two senses. The first positions ethnography as a counterpart to data, what Boellstorff has noted as the presentation of '"ethnography" as the Other to big data' (2013). From a position that takes data to be a competitor epistemology, ethnography is either at risk of losing its epistemic territory or an under-appreciated corrective that can balance out the reductionism of large-scale, quantitative methods. A now familiar critical stance is that big data simply misses out all that is human, and embodied, about living in the world. For example, the tech ethnographer Tricia Wang has argued, within the corporate context, that 'big data needs thick [ethnographic] data' (2016), drawing on conventional arguments in favor of mixed-methods research (cf. Seaver 2015).

The second approach finds anthropologists attending ethnographically to data practitioners themselves. We can already point to detailed analyses of the work and worlds of data scientists (Lowrie 2018; Williams 2018) Quantified Self trackers (Schüll and Ruckenstein 2017; Lupton 2013; Nafus and Sherman 2014) border management regimes (M'charek et al. 2014) and global health experts (Biruk 2018). With these ethnographies come methodological challenges. Taking their first steps into the field, ethnographers find that their field sites are often distributed, access is ethically fraught, and the technical legibility of practices opaque (Seaver 2017; Knox and Nafus 2019). Nonetheless, field sites for those who seek to study communities and practices data continue to multiply around the world: Quantified Self communities, data mining start-ups, clandestine data centres, biometric infrastructures, and more.

Many other contemporary ethnographers may be more accurately described as accidental ethnographers of data. They do not set out to study data, but rather find it central in the lives of the people with whom they work. To study the nation state, health systems, judicial systems, the economy, or scientific communities, is to come up against infrastructures, practices and discourses of data. A study of indigenous kinship might lead one to a genetic database, and from there into issues of sovereignty, ownership, and biocapital (TallBear 2013; Reardon 2017). An analysis of courtrooms might lead to software developers developing predictions of recidivism or newly spatialized carceral imaginations (Benjamin 2016, 2019). These accidental ethnographers of data will find it tightly knitted to core anthropological themes: kinship, economy, religion, law, and so on. Data cannot be set aside, because it is bound up with the development of new norms, with changing ideas of personhood, place, kin and more. Encountering data requires anthropologists to revisit their theoretical commitments.

Data as Theory

The rise of data has troubled settled agreements and re-opened normative discussion about matters of longstanding anthropological interest. When people turn to data to know themselves and their world, we find the meaning of social concepts changing. To make visible some of these transformations, we draw a parallel between the current data moment and an earlier moment where new technologies opened up and challenged the meaning of core anthropological concerns. The technologies and discussion around New Reproductive Technologies (NRTs) constitute a prior moment of transformative technological change which gave rise to broad anthropological re-theorizing. By reading through that moment, we set out to identify similar themes critical to the current moment.

A brief recapitulation of NRTs – both their technical capacities and the resulting work in anthropology – is useful. As the now familiar account goes, in the early 1990s, the development of in vitro fertilization techniques led to new possibilities for human procreation, and reopened public discussions about how kin could be conceptualized. This sudden prominence of NRTs in Euroamerican consciousness demanded that the public – and anthropologists – reconsider a

fundamental and take-for-granted aspect of life: reproduction. From conception to parenthood, in vitro new technologies challenged understandings of the artificial and the natural, and assumptions about what is available for human intervention (Wagner 2016 [1975]; Wagner 1977); in so doing, they undid the very grounds upon which such categories organized and shaped the world in the first place (Strathern 1992a). As legal cases made their way through different national courts, and teams of lawyers debated the new meanings of parent and child, in anthropology, two moves followed.

First, new technologies meant new field sites: kinship was no longer the domain of the domestic, nor the way to characterize the exoticized other - it was in laboratories, hospital waiting rooms, and, more recently with the advent of consumer genetics, corporate board rooms (Franklin and McKinnon 2001; Reardon 2017). The entities that populated these field sites - such as cell lines, embryos, technologies, technical procedures - shifted in and out of existing categorizations, creating a need for conceptual re-theorization (e.g. Thomson 2007). Second, kinship was suddenly once again a lively vernacular social category, giving it new hold on the anthropological imagination (Franklin and McKinnon 2001: 173). A shared public and academic question emerged: how would people orient themselves to what seemed to be a tectonic shift in how they conceived life, and yet was, at the same time, uncannily familiar in the social forms it reproduced? (Franklin 2013).

Today, we see ‘data’ moving research similarly in both of these registers: it opens up field sites, and it prompts questions about what is new and what is not in the way data reproduces the social. While the former is reflected in ethnographies of data practice, the latter, more theoretical, concern is still emerging. Drawing from the legacies of NRTs, we reflect on what an anthropology of data which is both attentive to shifting tectonics and similarity of form, might do well to attend to. In what follows, we tack back and forth between ethnographic and theoretical work on data, and critical commentaries on NRTs, using the comparison to draw out categories troubled, presumptions that no longer hold, and novel contours to social life.

Calculative Compositions

Like NRTs, data collection opens up new aspects of life for intervention and manipulation. The body is an ideal example: iris scans, facial recognition, gait analysis, voiceprints, and even analprints (Park et al. 2020) make new markers of identity accessible and claim to surface deeper meanings about people. There is also the routine but voracious collection of mundane data online – data on everyday intimate, personal, private activity – framed as a source of valuable insight (Zuboff 2018; Couldry 2019). Where NRTs stressed the binary between the natural and artificial, these data practices draw into dispute the interior and exterior, the intimate and the productive, and the private and the public.

These new forms of visibility are also new arenas for calculation. The presence of data asks people to re-invest in the power of numbers (Porter 1995): the idea that nothing can escape quantification and ceaseless calculation.Consequently, the data moment also asks us to refocus our anthropological attention on social practices of quantification and measurement. While quantification and numbers have often been analysed (and thoroughly critiqued) as being seductive because they simplify and reduce social life (Engle Merry 2016), we often find data used to figure life as extremely complex. From the neuroscientists in Rayna Rapp's examination of big data neuroscience who end up looking for a 'needle in a haystack' (Sullivan 2013, cited in Rapp 2016: 8) to the government officials in Louise Amoore's accounts of new forms of data-driven securitisation who warn that there are untold threats "hiding in the data" (2006), big data is understood by many of its practitioners to replicate the messiness and complexity of life; renowned MIT computer scientist Alex Pentland calls his work with big data "reality mining" (Eagle and Pentland 2006). But we might say, as Franklin said of IVF, that data is both "like and unlike what it imitates" (2013: 8). The "insight" data gives works through the collapse of the tension between entities and their representation in data, such that its proponents can claim that 'your data' knows you better than you know yourself (Douglas-Jones, this volume). In so doing, however, it also makes 'yourself' newly available for re-working, re-assimilating, re-configuring, re-evaluating. "Deeper insight" is revealed as, in fact, dependent on creative acts that make people (or communities, or things) tractable to emergent forms of governance, from state interventions to grassroots appropriations to self-directed bodily optimisation (Nafus and Sherman 2014; see also

Nair this volume). Within the representational logic of many big data practices there are therefore simultaneous enactments of both conservation and transformation.This demands an equal sensitivity to the constitutive tension between these two qualities in our critical approaches .This is apparent in, for example, Simone Browne's analysis of biometric data collection, in which she takes Frantz Fanon's notion of 'epidermalization' employed to describe the 'marking of the racial Other' as a fractured 'body out of place' (Browne 2010: 134), and re-works it as 'digital epidermalisation', in order to take account of how the apparently disembodied and neutral technologies of contemporary biometric data collection operate to profoundly destabilise the ontological security of black bodies. Here, technologies that are claimed to be beyond race (ibid: 143) are revealed as the mechanisms of racialisation itself.

Thus the challenge arises: how to handle the tension between the claims made on behalf of data, and the realities of data's effects? This is one of the pressing questions for an anthropology of data. Scholars across disciplines have remarked on shifts or displacements in the idea of knowledge production at work in the way people make claims with data, particularly big data (Walford 2021), as for example when practitioners suggest that aggregate units such as "class" or "gender" are no longer relevant in the face of the hyper-granularity of big data (cf Cheney-Lippold 2011). Shifts such as these moves knowledge production into familiar but not-quite recognizable epistemological terrain. Natasha Dow Schüll notes that Quantified Self practitioners reach uneasily for terms such as "quantitative autobiography" (2019: 31) in order to capture the way their self-knowledge slips between fact and fiction. In a different context, Orit Halpern uses the term "communicative objectivity", to signal a "new aesthetic and practice of truth; a valorization of analysis and pattern seeking" (2014: 15) discernible in the construction of "smart" data-driven environments. Catelijne Coopman's description of data mining as a form of "artful revelation" (2014), plays exactly with the trope of hidden depths and visible surfaces experienced by analysts of data sets. Similarly, Louise Amoore characterises the practices employed by the US Government to "flag" the presence of terrorists 'in the data' as working through an "ontology of association" (2011:27), in which disparate data points are connected up temporarily in order to create a fleeting, but potentially life-changing, risk-based and often racialised identity. In these examples, key epistemological principles that have commonly been ascribed to scientific

knowledge – universality, replicability, objectivity, stability through time – no longer fully apply to the claims being made; and yet the resulting analyses are taken as truth.

While pattern analysis, temporary association, and revelation are familiar to Euroamerican sensibilities, they also lead to conclusions that are difficult to fully recognize, not just analytically, but in terms of their representational construction, legal status, social and ethical implications. In data's torqued extensions of what is already known, then, quantification and measurement are not just social practices, they are constitutive of specific forms of sociality that claim to reach beyond the social categories that social scientists are fluent in. Anthropologists of data need to develop a sensitivity then to both the claims made of data's capacity to re-shape the world, and also keep a firm grip on the tools at hand to anchor those claims in specific histories of practice and thought.

Retention and Reinvention of Form

Drawing on such histories will allow anthropologists to nuance the wider debate on data that often relies on an overly simplified relationship between the old and the new. Numbering and enumeration are powerful, and as we have mentioned, data belongs to calculative regimes that rest on prior histories of listing, counting, specifying (Porter 1995). The creation of data by nation states to tell stories about people and place are rightly analyzed and critiqued as today being in continuity with colonial violence (Couldry and Mejias 2018; Kukutai and Cormack 2019, Thatcher et al. 2016), retaining administrative legacies stark in national statistics (Ruppert and Isin 2019; cf Hull 2012). Biometrics, Simone Browne points out forcefully, is a direct descendent of Victorian practices of anthropometry and, before that, the practice of branding slaves which coerced the body's surface into a determination of their identity as property (Browne 2015, see also Chun 2009). But these are not the only histories to tell of data. Halpern argues that it makes little sense to consider imaginaries of data-driven self-regulating 'smart cities' without simultaneously drawing links to cybernetic imaginaries that accompanied earlier eras of computational dreaming (Halpern 2014). Anthropologists need to look to such histories; but they also need to ethnographically trace out the means by which the contemporary data moment is re-working and re-shaping such histories in ways that have yet to be fully understood.

Here, we take inspiration from critical archivist Ann Laura Stoler, who asks how imperial legacies of inequality endure so recognisably and yet, at the same time, so invisibly - not as "mimetic versions of earlier imperial incarnations" but "refashioned" to be "ineffably threaded through the fabric of contemporary life forms [such that] they seem indiscernible as distinct effects, as if everywhere and no-where at all" (2016: 4) For example, if the legacies of quantification, and the concomitant 'birth of statistics', can be credited with inventing the idea of the "normal" (Hacking 1990; cf Foucault 1977 [1975]) - and of course, the "abnormal" - which has subsequently woven itself into the way many people now conceptualise the world, then the turn to big data might be seen as permitting the exploration of other charismatic and emergent forms of social distribution, perhaps along the lines of what Michelle Murphy calls 'phantasmagrams … intangible form[s] brought into sensibility as a palpable presence with the help of quantitative practices' (Murphy 2015, np, see also Murphy 2017). Murphy is here thinking specifically of economic forms, but we might also think of the way recommendation algorithms draw on vast amounts of consumer data to re-distribute people according to logics of 'likeness' and 'liking', for example (Seaver 2012; Lury and Day 2019). Whilst there is no doubt that contemporary data practices are forms of "infopower" (Koopman 2019), it is an open ethnographic question as to what specific shapes these forms are taking.

As Minna Ruckenstein and Natasha Dow Schüll argue, ethnography here becomes a means to negotiate the complex power dynamics that are woven through datafication, 'by revealing how data and its technologies are taken up, enacted, and sometimes repurposed' (Ruckenstein and Schüll 2017: 265). Whilst contemporary data practices often present familiar power dynamics, people are also taking hold of data in projects of refusal and resistance, whether collecting Data for Black Lives or enlisting citizens in scientific initiatives (Watson-Daniels et al. 2020, Gabrys et al. 2016). As Dana Greenfield suggests, following Michelle Murphy's lead, self-tracking data practices might also be seen as a form of "counter conduct" (Murphy 2012), in the vein of the consciousness-raising activities of feminists in the 1960s and 1970s in which women appropriated speculums to do their own vaginal examinations (Greenfield 2016: 134). Stories told with and about data matter: histories of specific data forms, but also knowledge forms are being actively

drawn on and folded into contemporary data-driven practices and discourses. They produce specific configurations of resistance and assimilation.

Anthropology has much to offer current critical engagement with data, not least by decentring the focus on Euroamerica which at present dominates the scholarly literature on the subject. Established anthropological debates around personhood, relations, society, nature, the state, value are all valuable tools to theorise the emergent phenomenon of 'data' as we have described it. However, it is also clear that an anthropological engagement with data demands that anthropologists develop and modify these theoretical approaches in order to take account of the only-partially recognisable social and cultural forms that data practices are producing. This will require an openness to other disciplines as much as a consolidation and extension of our disciplinary conventions. Aptly, it is in this spirit of conservation and transformation that we approach an anthropology of data.

Papers in this Volume

The pieces collected here explore a range of possible directions for an anthropology of data that is attentive to historical continuities and disjunctions, and to the conditions for thought that the present data moment has engendered. They explore these broad concerns in particular locations, drawing out the significance of data for topics of longstanding anthropological concern. Here we introduce and gather them together in five thematic areas: aesthetics, temporality, economy, composition, and intensities.

Data Aesthetics

We begin with Vijayanka Nair's detailed ethnography of Aadhaar, the nation-wide biometric identity program of the Government of India. Tracing the project from its policy inception into registration offices and, eventually, to the courts, Nair observes a shift in the program's purpose. Early in its history, Aadhaar was framed as a tool to help citizens prove that they were indeed who

they said they were, merely registering an already existing social identity. Yet, as the system took shape, the program's central task turned from verifying 'are you who you say you are' to the much more vexed question of 'who are you?' Moving back and forth between the database and its reception, Nair finds that various technical aspects of Aadhaar, from the shape of the registration interface to the construction of the database, re-stage the broader tension between individuality and dividuality that Indian citizens have often found themselves caught in as they negotiate their political agency. Paying attention to the aesthetic 'surface' of the database – the literal, bureaucratic form through which data is made – Nair demonstrates that 'being translated into data is a complex, not to mention fraught, process' (Nair, this volume).

Contemporary data discourses often claim for data a privileged access to deeper social reality; attention to the process of 'becoming data', as Nair puts it, demonstrates the contingencies and flaws inherent in such 'dataism' (Van Dijck 2014). But there is more to data than representational adequacy, as Nair's ethnography demonstrates. Data's aesthetic form, whether in the design of interfaces or the shape of visualization, elicits social effects. In Nick Seaver's article, we encounter a form of data aesthetics endemic to contemporary machine learning: the representation of culture as a mathematized, mappable space. To make sense of contemporary music recommender systems, which analyze user interaction data to create 'music spaces' full of users and music to be recommended, Seaver tells the history of a related technique with roots in anthropology: multidimensional scaling. Tracing the interlinked history of anthropology and computing through the 1960s and 1970s, Seaver finds that spatializing practices are used by post-war cognitive anthropologists to grant elusive 'cultural' phenomena a sense of reality by borrowing from the world of physical objects and geographic distance. Spatializing data creates new surfaces, transforming sparse and discrete data points into apparently continuous cultural environments. Nair and Seaver's articles suggest the importance of a move away from a representational critique of data; their analyses demonstrate the effects that data has in the world irrespective of its accuracy. An aesthetic sensibility, and attention to the persuasiveness of form, opens up critical readings of data beyond representational paradigms.

Data Times

Several papers in this collection attend to the ways that data is caught up in temporalizing projects. In conventional critiques, data's apparent fixity contrasts with the dynamic world it is marshaled to represent. Anthropological approaches to data recognize, as Tom Boellstorff has argued, that 'data is always a temporal formation' (2013: no pagination) – it is *dated*, sampled from a particular moment in time. But anthropology can do more than re-locate data in its proper temporal context: we can examine how data is enlisted in efforts to produce new temporalities.

Data times are diverse, competing, and often contradictory temporal frames. In Tahani Nadim's analysis of the material stuff of natural history archives, she reads digitized and physical records of biological species together, finding the temporalities of both data collection and storage entangled with the history of colonialism. Nadim's paper expands our temporal horizon, spanning from early pen-and-paper days to new methods of DNA barcoding. Those new methods, like many data initiatives, distinguish themselves not only by their 'bigness' but their claims to speed. Advocates for new techniques bemoan the 'taxonomic impediment' – the slowness of conventional methods for collecting, analyzing, and classifying biodiversity data – and promise a future in which DNA sequencing data will provide a total archive of Earth's animals, plants, and fungi. The speediness of this new data is not only a resource, but also a demand: this digital salvage taxonomy requires urgent action, before the anticipated decimation of earthly biodiversity arrives. Salvage imagery appears elsewhere in the collection, when an interlocutor of Nair's suggests that data in the Aadhaar system offers the promise of immortality: absent any policy for deletion, data can 'live forever' (see also Taylor, this volume). The Digital India project, Nair tells us, explicitly aims to reset a colonial past with post-colonial sensitivities; the static archival imaginary is set against promissory visions of a digital future.

The relationship between data and imagined futures takes on a new concreteness in the 'data bunkers' described in A.R.E. Taylor's contribution: here, data is not fundamentally static and durable, but rather a vulnerable collection of networked traces, always at risk of being wiped out of existence by catastrophic events. Anxieties about future data loss manifest in a concern for the physical arrangement of information, the location of backups and the wires along which data flows, when every connection is a potential risk. These bunkers, often repurposed Cold War bomb shelters, reach forward and backward in time for their meaning, showing what happens when the

protection of data and its integrity displaces the now-dated image of the family hunkering down with canned food. Data times appear in myriad forms across these papers, unsettling accepted genealogies and taken-for-granted futures of the data present.

Data Economies

Given the entanglement of data and capital described above, it is not surprising to find concerns about economy threaded through the papers. In 2016 the Financial Accounting Standards Board declared that data should potentially be listed as a separate 'tangible corporate asset' on balance sheets, though as yet with no agreed means of calculating 'fair value' (Monga 2016). But across the papers in this collection, data is made valuable in diverse ways, entering economies not always recognizable as financial. Hannah Knox, in her article on the careful and hopeful acts of energy trackers, shows how the value of data is both about saving money on heating bills and about a new understanding of the home gained through working with the data sets over large coffee shop tables at meet-ups. At a moment when 'data-driven' decisions are commonly valorized, Knox's ethnography of environmental data presents the complex unfolding of this drive, as people work to make sense of data produced by sensors in their homes. A concern for data – and for giving data force – draws people into new relations, and data generates forms of relationality that challenge conventional ethnographic imaginaries. Knox puts forward the 'hack' as a method for anthropological engagement with data practices – one which takes data 'not just as a representation that we need to deconstruct, but as a means of engaging relations that are imprecise and unknown'. Situating anthropology in data streams and assemblies as people worry and wonder about their home's energy consumption, Knox asks: how can we think data with and through capitalism? Antonia Walford's article about environmental data collaborations in the Brazilian Amazon presents another setting where the value of data is multiple. The data economies of the Large Scale Biosphere Atmosphere project are those of scientific collaboration, belonging to complex relations and socialities. Walford's close attention to how data is made, and its capacities for transformation, gives us insight into its conceptual flexibility as it moves between people with different relations both to it and to each other. For some, data is valuable because it takes work to collect, but for others, it has value because it can be made into something else, used and used again in future research collaborations internationally, beyond Brazil. By foregrounding the labour involved in

the generation and reproduction of data's social value – including cleaning nuances of alienation, ownership and rights in and for data, Walford's analysis demonstrates that to consider data as merely a commodity or currency is to overdetermine what else it might be. These details offer ethnographic differentiation to imaginaries of a global data economy, and provide a counterpoint to the clean association of data with financial value.

Data Compositions

Nair and Knox's papers illustrate how data imaginaries can serve to compose and recompose 'publics' in different forms, through the centralized operations of a government database or the distributed interpretive work of local enthusiasts. Rachel Douglas-Jones' paper zooms in on the act of composition itself., tracing how 'bodies' of data come to refer both to physical bodies and to bodies politic; examining how data is thought of as a means to knowledge about such bodies. How might data collection be shaped to invoke new knowledges, and to what ends? Douglas-Jones begins with data on and for the physical body, looking at Scandinavian communities convened around the use of everyday tracking devices. She examines the cultural understandings of the body that underly efforts to compose sovereign individuals who are coextensive with their 'personal' data. Moving from the sovereign body to the sovereign people, she shifts the site of analysis to Indigenous Data Sovereignty movements. Here, what data is, and is about, becomes a matter of how stories are told and how prior modes of collecting data (though government statistics) give shape to what data will be, and how those peoples will be known.

Such data analytic practices appear to validate long-held anthropological ideas about persons and groups: datafied persons are conspicuously partible, decomposed into collections of interests for the purposes of advertising and re-aggregated into countless new collectives along axes of partial connection. But the power to aggregate and disaggregate – to classify or to individuate – is not evenly distributed. These questions of the power to define come forward in Sarah Blacker's account of a report on the environmental and health effects of toxins flowing through Lake Athabasca in Alberta, Canada. The object at the centre of Blacker's analysis is a technique known as the 'three track methodology', which was designed to allow the inclusion and recognition of indigenous knowledge within official policy making. . Put to use under fraught political conditions, the methodology represents an attempt to compose data in such a way that it

can be legible to western science, recognizable as policy relevant. Blacker analyses how different kinds of knowledge are made into data, working through the aesthetic, temporal and incommensurable dimensions present in the making of contamination evidence. Her article suggests that data cannot be read without its national and historical framing – framings which become part of the question 'what is data'?

In both Blacker and Douglas-Jones's articles, data shows its capacity for composition, bringing persons and data into new social and political formations. Both accounts refuse the spatialization of a technological timeline (Fabian 1983) when considering the uses to which indigenous knowledge and data will be put. An anthropology of data works towards developing theory to accommodate data's compositional capacity, and, we suggest, contributes to re-composing anthropological attention at the same time.

Cori Hayden's contribution turns to another form of composition, analyzing how concerns about connection and contagion from turn-of-the-century crowd theory have become newly relevant with the emergence of online 'crowds' that are at once the subjects and objects of massive data production. As she writes, 'data – in its form of clicks, and shares, and micro-targeting, and re-tweets, and virality, and filter bubbles – has been a source of, and a site of anxious concern about, emotional contagion, social "dissolution", more-than-human social compositions, and pressing questions about the mediums through which ineffable energetic forces travel' (Hayden, this volume). Where the power of 'big data' is conventionally associated with its size and scope, Hayden traces an argument for attending to the modes of connection within the crowd, the shape of the network through which social intensities flow and act. Through a careful reappraisal of work by notorious theorists like Gustave Le Bon, Hayden demonstrates the value of expanding our theoretical frames of reference when working to understand such apparently novel dynamics as social action in online spaces. The presence, availability and behaviour of data in these spaces prompts the revisitation of questions central to the crowd theory of the nineteenth century.

Conclusion

Data is generated for and drawn into existing social worlds and problems. In data practices, we recognize continuities with deeper pasts, with projects of society making and attempts at 'managing' people and their worlds. How it will come to speak to and shape those worlds and problems cannot be known in advance. The anthropological capacity to critically regard the claims of the new, and to see how old ideas appear in new guises is necessary at a juncture in great need of highlighting new injustices in the name of new freedoms.

As a powerful generic, linking widely varying practices and objects to each other through its apparently modest epistemic form, an anthropology of data should, following Strathern (2014), recover the specifics of this generic, locating 'data' in its contexts. But, as papers in this collection demonstrate, it remains necessary and generative to think data across sites. From the informatic rendering of genomics (Tutton and Prainsack 2011) to the citizen as a digitally re-articulated configuration of domination and resistance (Ruppert 2012; Lyon 2008), or the vast Earth BioGenome project as systems of exchange flourishing through data, social practices are valued on data markets that exist at the edges of conventional regulatory apparatuses (Gerlitz and Helmond 2013; Maurer 2015). Much is being worked out. These studies, and many others like them, demonstrate the immense intellectual potential of the social study of data. We close with call for further work that thinks across sites and settings to produce concrete theoretical formulations. An anthropology of data provides ethnographic thickness and sited-ness to counter data's ideologies of objectivity; it can also open up new conceptual approaches for thinking with and about social worlds, as they are used, made, and done through data. For data makes relations, and it is the careful analysis of the consequences of this capacity to mold and re-shape hoped-for futures towards which an anthropology of data should direct itself.

Acknowledgements

The authors thank the reviewers of this collection. We learned after the fact that the late Sally Engle Merry had provided insightful comments which helped us reshape our arguments. Thanks are due also to Katie Smith and Nayanika Mathur for their critical readings of work in progress, and to Jessica Turner and Lara Tatiana Reime for their support of the editorial process. We thank

the participants of the seminar that gave rise to this collection, in particular discussants Haidy Geismar and Marilyn Strathern.

References

Amoore, L. 2006. Biometric Borders: Governing mobilities in the war on terror. Political Geography 25: 336-351

Amoore, L. 2011. Data Derivatives: On the Emergence of a Security risk Calculus for Our times. Theory, Culture and Society 28(6): 24–43.

Benjamin, R. 2016. Catching our breath: critical race STS and the Carceral Imagination. Engaging Science, Technology and Society 2: 145-156.

Benjamin, R. (ed). 2020. Captivating Technology: Race, Carceral Technoscience and Liberatory Imagination in Everyday Life. Durham, NC: Duke University Press.

Boellstoff, T. 2013. Making data big, in theory. First Monday 18(10).

Browne, S. 2010. Digital Epidermalization: Race, Identity and Biometrics. Critical Sociology 36(1): 131–150.

Browne, S. 2015 Dark Matters: On the Surveillance of Blackness. Durham, NC: Duke University Press.

Cheney-Lippold, J. 2011.A New Algorithmic Identity : Soft Biopolitics and the Modulation of Control. Theory Culture Society 28(6): 164–181.

Chun, W. H. K. 2016. Programmed Visions: Software andMemory. Cambridge, MA: MIT Press.

Chun, W. H. K. Introduction: Race and/as Technology; or, How to Do Things to Race (eds.) L. Joyrich & W.Chun Camera Obscura 70, 42(1): 7–35.

Couldry, N. and U.Mejias. 2018. Data colonialism: rethinking big data's relation to the contemporary subject. Television and New Media.

Crowder, J. W., M. Fortun, R. Besara, &L. Poirier. 2019. Anthropological Data in the Digital Age: New Possibilities, New Challenges. London: Palgrave Macmillan.

Dalton, C. & J. Thatcher. 2014. What does a critical data studies look like ,and why do we care? Seven points for a critical approach to 'Big Data' Society & Space.

Dourish, P. 2017. The Stuff of Bits: An Essay on the Materialities of Information. Cambridge: MIT Press.

Dourish, P., & Mainwaring, S. (2012). Ubicomp's colonial impulse. Proceedings of UbiComp '12 (pp. 133– 142).

Eagle, N & A. Pentland. 2006. Reality mining: sensing complex social systems. Pers Ubiquit Comput 10: 255–268.

Engle Merry, S. 2016. The Seductions of Quantification. Chicago: University of Chicago Press.

Engelke, M. 2008. The Objects of Evidence: Anthropological Approaches to the Objects of Knowledge' Journal oftheRoyal Anthropological Instituten14(1): sii–S158.

Fabian, J. 1983. Time and the Other: How Anthropology Makes its Object. New York: Columbia University Press.

Foucault, M. 1977 [1975]. Discipline and Punish, Alan Sheridan (trans.), New York: Pantheon Books.

Fourcade, M. & K. Healy. 2017. Seeing like a market. Socio-Economic Review 15(1): 9–29.

Franklin, S. 2003. Rethinking nature-culture: anthropology and the new genetics. Anthropological Theory 2003 3(1): 65–85.

Franklin, S. 2013. Biological Relatives: IVF, Stem Cells, and the Future of Kinship. Durham, NC: Duke University Press.

Franklin, S. and McKinnon, S. 2001. Relative Values: Reconfiguring Kinship Studies. Durham, NC: Duke University Press.

Gabrys, J., H. Pritchard and B.Barratt. 2016. Just good enough data: Figuring data citizenships through air pollution sensing and data stories. Big Data and Soc July–December 2016: 1–14.

Gardiner, E. and Musto, R.G. 2015. The digital humanities: a primer for students and scholars. Cambridge: Cambridge University Press.

Gerlitz, C. & Helmond A. 2013. The Like Economy: social buttons and the data-intensive web New Media & Society 15(8), 1348–1365.

Greenfield, D. 2016. Deep data: Notes on the n of 1. In D. Nafus (Ed.), Quantified: Biosensing technologies in everyday life, 123–146.

Hacking, I. 1990. The Taming of Chance. Cambridge: Cambridge University Press.

Hales, M. K. 2019. Animating Relations: Digitally Mediated Intimacies between the Living and the Dead. Cultural Anthropology 34(2): 187–212.

Halpern, O. 2014. Beautiful Data: a history of vision and reason since 1945. Durham, NC: Duke University Press.

Hastrup, K.2004. Getting it right: Knowledge and evidence in anthropology. Anthropological Theory 4(4):455–472.

Hull, M. 2012. Government of Paper: The Materiality of Bureaucracy in Urban Pakistan. Berkeley: University of California Press.

India Today. 2019. Ghosts didn't vote in Lok Sabha polls, all were humans, says EC rubbishing claims on data mismatch. India Today June 1st2019.
(Available online: https://www.indiatoday.in/india/story/ghosts-did-not-vote-2019-lok-sabha-polls-data-mismatch-election-commission-1540291-2019-06-01 Accessed June 1st 2019.)

Kitchin, R. & Lauriault T.P. 2014. Towards critical data studies: Charting and unpacking data assemblages and their work. The Programmable City Working Paper 2. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Kitchin, R. 2013. Big Data and Human Geography: Opportunities, Challenges and Risks. Dialogues in Human Geography 3, 262–67.

Knox, H. & D. Nafus. 2018. Ethnography for a Data Saturated World. Manchester: Manchester University Press.

Kukutai, T. and D. Cormack.2019. Mana motuhake ā-raraunga: datafication and social science research in Aotearoa, Kōtuitui: New Zealand Journal of Social Sciences Online.

Lowrie, I. 2018. Becoming a Real Data Scientist: Expertise, Flexibility and Lifelong Learning. In Ethnography for a Data Saturated World (eds.) H. Knox & D. Nafus, 62-81. Manchester: Manchester University Press.

Lupton, D. 2013. Quantifying the Body: monitoring and measuring health in the age of health technologies. Critical Public Health 23 (4), 393-403.

Lury, C. and S. Day. 2019. Algorithmic Personalization as a Mode of Individualization.Theory, Culture and Society 36(2): 17–37.

Lyon, D.2008. Biometrics, identification and surveillance. Bioethics 22 (9), 499-508.

M'charek, A., K. Schramm & D. Skinner. 2014. Topologies of Race: Doing territory, population and identity in Europe. Science Technology and Human Values 39 (4), 468-487.

Marres, N. (ed). 2017. Digital Sociology: The Reinvention of Social Research. Wiley.

Maurer, B. 1997. Recharting the Carribean: Land, Law and Citizenship in the British Virgin Islands. Ann Arbor: University of Michigan Press.

Maurer, B. 2015. Principles of Descent and Alliance for Big Data. In Data: Now bigger and better! (eds.) T. Boellstorff & B. Maurer. Prickly Paradigm Press.

Merry, S.E. 2016. Seductions of Quantification: Measuring Human Rights, Gender Violence and Sex Trafficking. Chicago: The University of Chicago Press.

Monga, V. 2016. Accounting's 21st Century Challenge: How to Value Intangible Assets? The Wall Street Journal March 21, 2016.
(Available online: https://www.wsj.com/articles/accountings-21st-century-challenge-how-to-value-intangible-assets-1458605126)

Mullaney, T. (ed.). 2019a. The Chinese Deathscape: Grave Reform in Modern China. Stanford: Stanford University Press.
——— 2019b. No room for the dead: on grave relocation in contemporary China.
(Available online: DOI: 10.21627/2019cd/tsm)

Murphy, M.2015. Phantasmagrams of Population. Histories of the Future. http://histscifi.com/essays/murphy/phantasmagrams-of-population Accessed 15th May 2020

Murphy, M. 2017. The Economization of Life. Durham, NC: Duke University Press.

Nafus, D. & J. Sherman. 2014. Big Data, Big Questions: This One Does Not Go Up To 11: The Quantified Self Movement as an Alternative Big Data Practice. International Journal Of Communication 8 (11), 1784–1794.

Nafus, D. 2018. Working ethnographically with sensor data. In Ethnography for a Data Saturated World (eds.) H. Knox & D. Nafus, 233-251. Manchester: Manchester University Press.

Nelson, D.M. 2015. Who Counts? The Mathematics of Death and Life after Genocide. Durham, NC: Duke University Press.

Park, S., D. D. Won, B. J. Lee. et al. 2020. A mountable toilet system for personalized health monitoring via the analysis of excreta. Nature Biomedical Engineering 4: 624–635.

Pfaffenberger, B. 1988. The Social Meaning of the Personal Computer: or, Why the Personal Computer Revolution Was No Revolution. Anthropological Quarterly 61(1): 39–47.

Porter, T. 1995. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton, NJ: Princeton University Press.

Power, M.1997. The Audit Society: Rituals of Verification. Oxford: Oxford University Press.

Rapp, R. 2016. Big data, small kids: Medico-scientific, familial and advocacy visions of human brains. BioSocieties 11: 1–21.

Reardon, J. 2017. The Post-Genomic Condition. Chicago: University of Chicago Press.

Reeves, M. 2016. The Black List: On Infrastructural Indeterminacy and its Reverberations. In Infrastrucutres and Social Complexity (eds.) Bruun Jensen, C. P. Harvey & A. Morita, pp. 296–308. London: Routledge.

Riles, A. (ed.) 2006. Documents: Artifacts of Modern Knowledge. Ann Arbor: The University of Michigan Press.

Ruckenstein, M. & N.D. Schüll. 2017. The Datafication of Health. Annual Review of Anthropology 46, 261-278.

Ruppert, E. 2012. The governmental topologies of database devices. Theory, Culture & Society 29, 4–5, 116-136.

Isin, E. and E. Ruppert.2019. Data's empire: postcolonial data politics. In Data Politics: Worlds, Subjects, Rights, (eds). D. Bigo, E. Isin, & E. Ruppert. London: Routledge.

Sandbukt, S. 2018. The Future of Digital Payments: The impact of disintermediated and decentralised financial technologies in Indonesia. Work in Progress paper, IT University of Copenhagen.

Savage, M. & R. Burrows. 2007. The Coming Crisis of Empirical Sociology. Sociology 41(5), 885–99.

Seaver, N. 2012. Algorithmic Recommendations and Synaptic Functions. Limn 2 Crowds and Clouds. https://limn.it/articles/algorithmic-recommendations-and-synaptic-functions/

Seaver, N. 2015. Bastard Algebra. In Data: Now bigger and better! (eds.) T. Boellstorff & B. Maurer. Prickly Paradigm Press.

——— 2017. Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. Big Data & Society 4(2), 1-12.

Star, S. and K. Ruhleder. 1996. Steps Toward an Ecology of Infrastructure:Design and Access for Large Information Spaces. Information Systems Research 7(1): 111-134.

Strathern, M. 2005 [1991] Partial Connections. Oxford: Altamira Press.

Strathern, M. 1992a. Reproducing the future: essays on anthropology, kinship and the new reproductive technologies. Manchester: Manchester University Press.

Strathern, M. 1992b. After Nature: English Kinship in the Late Twentieth Century. Cambridge, Cambridge University Press.

Strathern, M. 1992c. Parts and wholes: refiguring relationships in a post-plural world
In Conceptualizing Society, (ed).A. Kuper.London: Routledge

——— (ed.) 2000. Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy. Routledge.

Street, A. and A. Kelly. 2020. Counting coronavirus: delivering diagnostic certainty in a global emergency. March 6, 2020, Somatosphere http://somatosphere.net/forumpost/counting-coronavirus-diagnostic-certainty-global-emergency/

Terras, M., J. Nyhan & E. Vanhoutte (eds). 2013. Defining Digital Humanities: A Reader. Farnham, England: Ashgate Publishing.

Thatcher, J., D. O'Sullivan & D. Mahmoudi. 2016. Data colonialism through accumulation by dispossession: New metaphors for daily data. Environment and Planning D: Society and Space 34(6): 990–1006

Thomson, C. 2007. Making Parents: The Ontological Choreography of Reproductive Technologies. Cambridge, MA: MIT Press.

Tinati, R., S. Halford, L. Carr, & C. Pope. 2014. Big Data: Methodological Challenges and Approaches for Sociological Analysis. Sociology 48(4), 663–81.

Tutton, R. & Prainsack, B. 2011. Enterprising or altruistic selves? Making up research subjects in genetic research. Sociology of Health and Illness 33 (7), 1081-1095

Verran, H. 2010. Number as an inventive frontier in knowing and working Australia's water resources. Anthropological Theory 10(1): 1–8.

Watson-Daniels, J, Y. Milner, N. Triplett, I. Headen, D. Day, Z. Bailey, M. Styles, L. Clinton, C. Andrews, M. Wilson, N. Ezeokoli, S. Jebbett Bullard, L. Mason-Brown. 2020. *Data for Black Lives COVID-19 Movement Pulsecheck and Roundtable Report*. April 2020, http://d4bl.org/report

Wagner, R. 2016 [1977]. The Invention of Culture. Chicago: The University of Chicago Press.

Wagner, R 1977. Scientific and Indigenous Papuan Conceptualizations of the Innate", in Bayliss-Smith, T. and R. Feachem (eds) Subsistence and Survival. New York: Academic Press, pp.

Wang, T. 2016. Why Big Data Needs Thick Data.
(Available online: https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7)

Williams, K. 2018. Engineering Ethnography. In Ethnography for a Data Saturated World (eds.) H. Knox & D. Nafus, 82-104. Manchester: Manchester University Press.

Zuboff, S. 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. London: Profile.

[i] For work that directly draws such comparisons between contemporary datafication and colonialism, see Couldry and Mejias 2019; Dourish and Mainwaring 2012; Ruppert and Isin 2019.

[ii] Out of cross disciplinary engagements, the reflexive field of Critical Data Studies has emerged as a transdisciplinary formation (Dalton and Thatcher 2014; Kitchin and Lauriault 2014).