Original Research Article

Datafication and data fiction: Narrating data and narrating with data

Paul Dourish[1] and Edgar Gómez Cruz[2]

Abstract
Data do not speak for themselves. Data must be narrated—put to work in particular contexts, sunk into narratives that give them shape and meaning, and mobilized as part of broader processes of interpretation and meaning-making. We examine these processes through the lens of ethnographic practice and, in particular, ethnography's attention to narrative processes. We draw on a particular case in which digital data must be animated and narrated by different groups in order to examine broader questions of how we might come to understand data ethnographically.

Keywords
Ethnography, datafication, narrative, trajectory, temporality

Introduction

Whatever their sources—sensor streams or written records, scientific instruments or ethnographic observations—data do not stand alone. They do their work in relation to multiple other entities. First, they do their work in relation to other data and to other data sets, through many different sorts of relations—providing supporting or countervailing evidence, for example, through massification, by means of aggregation, or in the nature of their singular difference. Second, they do their work with respect to systems of processing—computers, databases, programs, algorithms, formulae, procedures, classifications, and counts. Third and perhaps most importantly, though, they do their work in relation to people. They frame new understandings, reinforce assumptions or experiences, decenter expectations, challenge dominant narratives, reveal phenomena, hide problems, and justify decisions. In this last relation, though, there remains a critical mediator. Data tells stories in the ways in which it is animated, explained, offered, and shared (Gabrys et al., 2016; Sharon and Zandbergen, 2017). Papacharissi (2015) uses the term "digital orality" to describe the ways in which data is embedded in narratives that produce situated knowledge. The ways in which data works often depend crucially on the way in which they can be narrated, and the way that these

narratives deal with what Pink et al. (2018) describe as "incomplete, contingent and fractured character of digital data". As Veel (2018) importantly notes, the ongoing entwining of data and narrative does not only constitute a reframing of data, but is also part of a reformulation of narrative and narrative practice.

The material presented here reflects discussions among a group of ethnographic researchers who came together at a series of workshops at RMIT University in Melbourne to discuss contemporary topics around data from a distinctly ethnographic perspective.[1] We came together not so much to share ethnographic accounts of data and data work, but rather, to ask what ethnography as a mode of inquiry might teach us about contemporary interests in data-driven scholarship and practice. As ethnographic researchers, we recognize that questions of perspective and voice are central to our methodology and to how ethnographic results are put to use. Our goal here is to draw on that

[1]Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California, USA
[2]School of the Arts and Media, University of New South Wales, Australia

Corresponding author:
Paul Dourish, Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA 92697-3440, USA.
Email: jpd@ics.uci.edu

experience in order to interrogate questions of how data speaks. Accordingly, we take up the questions of the relationship between data and narrative with an understanding that the two are deeply mutually entwined. We give no priority to one topic or the other, neither seeing data as inherently beholden to narrative nor narrative as entirely bounded by data. Indeed, we see neither data nor narrative as pre-given. Bearing in mind Strathern's (2003) observation that ethnography is a method for generating more data than the ethnographer is aware of at the time, we recognize that the issue of how things become data or are taken to be data is itself fraught and complex.

We understand "data" broadly here. As ethnographic researchers, our own data is of disparate types; not just notes, transcripts, and observations, but jottings, artifacts, feelings, and experiences; as Ortner (2006) comments, ethnography is a practice that uses "the self…as the primary instrument of knowing". We speak here also to data in other forms, including the sensor data and large-scale quantitative foundations of typical data analytics efforts. van Dijck (2014) borrows the term "datafication" from Mayer-Schoenberger and Cukier (2013) to refer to "the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis." (p. 198). We use the term here in a related manner, focusing not so much on the production of quantified data, but on the processes of symbolic and imaginative work that underlie coming to think of something as "data" in the first place. We use it too, of course, for its resonant play with the term "data fiction" (Nadim, 2016), highlighting not just the inevitable partiality of data but also its purposive role.

We think of the relationship between data and narrative in terms of two "scalar moves" in Big Data settings. The first move is the move from small to large, or from datum to data set, the massification involved in the collation of large collections of information. This move depends on logics of equivalence, and the claim that these data are sufficiently "alike" as to be able to be combined, compared, added, and divided, as exemplars or instances of a singular phenomenon, claims that have been productively examined in a range of domains by Martin and Lynch (2009). The second move is the move from large to small implicit in the drawing of conclusions or categories from data analysis. This move, by contrast, depends on the logic of correspondence; it rests upon the auspices by which we might say that a feature in the data corresponds to a feature in the world—a class of consumer, a type of event, or an item of interest in the domain about which the data "speak".

In both of these moves, acts of narration are key; both logics are narrative resources of the sort that ethnography often takes as central (Boellstorff, 2013). Our purpose here is to find these narrative acts at work and to examine some of the consequences and limits of narrative within the broader processes of data-driven analysis.

An anchoring case

To contextualize this discussion, we offer one example from prior work, discussed in greater detail elsewhere (see Shklovski et al., 2009, 2015; Troshynski et al., 2008). This case centers on efforts made in the mid-2000s by the California Department of Corrections and Rehabilitation to develop and test a system for the continual monitoring of the location and movements of paroled sex offenders.

Responding to legislative moves elsewhere in the United States, the department was interested in whether such a system was both technologically and organizationally feasible. Technological feasibility concerned questions such as battery life, accuracy, reliability, and the cost of tracking technologies. Organizational feasibility, on the other hand, turned on questions such as the organization's ability to manage, store, and work with the data that such a system would generate.

When released on parole, parolees are subject to a series of parole conditions, violation of which may cause them to be incarcerated again. These might include participation in group therapy processes and other rehabilitative schemes. Further, a series of spatial limits are typically placed on their movements, requiring that they stay within a range of, say, 25 miles of their place of residence, and excluding them from zones of 2000 feet around public parks, playgrounds, swimming pools, libraries, and schools. Conditions of this sort were laid down even before the deployment of GPS technology as part of parole, but GPS was viewed as a mechanism by which these conditions could be monitored and enforced. Given that the GPS tracking device itself then enters into the parole process, parole conditions require parolees to ensure that the device is not removed and that it is maintained in an operable state (e.g. kept charged.)

We[2] studied the process of conducting this evaluation from two perspectives, that of the parolees themselves, and that of the parole officers who managed their cases. In each case, we found narrative central to the ways in which they managed, oriented towards, and understood the data generated by the system in relation to their own experience.

From the perspective of the parolees, we will draw attention briefly here to two particular issues. The first was the opportunity to use the data in order to frame alternative narratives of their own movements

and habits. That is, repeat offenders who had previously spent time in prison and outside were familiar with the way in which they might be subject to police "hassles" and often targeted for attention when incidents, scares, or suspicions surfaced about sexual predation in the neighborhoods in which they lived. On these occasions, they might find themselves needing to be able to account for their movements or generate evidence about their presence and absence from particular places (in order, for instance, to signal that they had not been at the location of a particular incident that had taken place. Such evidence was often difficult to produce in usual circumstances, but the GPS monitor that was now attached to their ankle provided to the parolees a technological source of evidence of their own innocence in the face of these hassles. They could now point out to police officers that their location was recorded and available directly to criminal justice authorities, and that this record would show that they had not been present at the scene of an incident. The fact that the data record was not self-generated but rather data of their movements authorized by the state made such "testimony" more reliable than that of friends, workmates, or neighbors, while at the same time saving the parolees the potential embarrassment of having to explain to friends, workmates, or neighbors why such evidence might be needed. In other words, the location data that was being collected offered an alternative narrative of their movements to the potential narrative being explored by police officers, and helped the parolees to counter those narratives.

At the same time, the fact of the data—the presence of the device, the condition of monitoring, and the presence of location data as a basis of individual's engagements with the authorities—also itself supported particular narratives of identity. That is, for repeat offenders who were subject to the parole program and participated in the trial, the permanent presence of the tracking device strapped to their body reinforced an account of them as "sex offenders". It was a reminder of prior crimes and reinforced a notion that they had come to adopt of their crimes as a key element of their identity. They referred to it as a reminder of who they were and a constant presence that helped them guard against possible future infractions or the conditions that might lead to recidivism. This aspect of their experience became particularly marked in our data because, during the course of our study, a new law in California passed that mandated lifetime, real-time location monitoring of sex offenders. Where, in the initial stages of our study, we had worked primarily with repeat offenders who, on leaving prison, had opted to participate in this program, the subject population for the later stages of the study, after the new law had passed, was markedly different, including a number of

people for whom "sex offender" was by no means part of their self-image. These people, for instance, might never have served time in prison, but who had an old conviction for a minor sex offence (such as public lewdness or nudity, which might in some circumstances arise from public urination) and who now found themselves subject to a law that they felt was really designed for someone else. This group also found that the presence of the device—for themselves as much as for anyone else—re-narrated their identity in ways that they strongly rejected but could not ignore. Perhaps more importantly, it embedded them within a prior, pre-established narrative that seemed to close off possibilities for the ways in which they might be able to present themselves to others and navigate social encounters.

Questions of narrative, and the processes by which data could be placed in context in order to become meaningful, also manifested itself regularly in the work of the parole officers, who managed the data streams being generated by the tracking devices. Again, we focus on just two of these here, and refer the interested reader to more detailed elaborations in prior publications.

As evidenced by the nature of the parole conditions, detailed above, space, spatiality, and movement is an important element in the logic of parole enforcement for sex offenders. Questions of the places that people might be, what else might be going on there, with whom they might interact, and so on, were central elements of how parole officers think about cases and about parolees. Accordingly, the simple information that the GPS units might provide—essentially latitude and longitude—needed to be transformed, for the parole officers, into signals of "good" and "bad" places—the sorts of places where someone might be and the sorts of places that they should avoid. Similarly, patterns of movement mattered; deviations from the norm, systematic patterns of "hanging out" in one sort of place or another, and the broader implications of such patterns were also of great interest. Consequently, the parole officers found themselves needing to narrate the movements of parolees in their case-load in different ways; as people "on the right path," as people who might be exposed to dangers or problems, as people who were in stable patterns or "on the slide". The problem for parole officers, who were now encouraged to replace "eyes-on" surveillance with digital, is that it was difficult for them to determine what sorts of places were showing up on the map. They might know some places, but specifics are important (for instance, distinguishing whether the subject is in a hardware store, or in the bar next door). Parole officers found themselves needing to take the logs of someone's movements and visit those places in order to establish real-world correlates to the data in the system, and in order to figure

out what story someone's presence in a place might tell, or how their presence could be slotted into a series of conventional narratives about the experiences of parole and post-prison life.

Finally, here, the other key act of narration in which parole officers engaged was one of their own actions and their sense of responsibility. Especially in the later part of our studies, after mandatory monitoring had been legislated, parole officers found themselves managing their cases essentially through the lens of data, rather than through the combination of regular meetings and surveillance that had earlier characterized their work. What is more, the granularity of the data, both spatially (the geographical resolution that it offered) and temporally (the frequency of tracking signals) radically reconfigured the work that they had to do. Prior to the technology deployment, a parolee might have a parole condition that required them to stay more than 2000 feet from a school, for example; after the deployment, on the other hand, any occasion on which a parolee was, say, 1993 feet from a school was digitally signaled and needed to be accounted for. Since it would clearly be politically unacceptable for there to be signals on a digital trace that were left unexamined in the context of some later criminal act, all failures to meet parole conditions needed to be investigated and accounted for within the system. Much of this work, needless to say, was relatively insignificant—the result of walking down the wrong side of the road, or taking a bus through an unfamiliar neighborhood. Nonetheless, the very fact of a digital trace produced the necessity of an account. Parole officers talked about this in terms of the responsibilities that they took on, and which they could effectively discharge. They felt a responsibility towards the parolees whom they supervised, but found that the reduced amount of time that they now spent with them made it harder to fulfill those obligations. They felt a responsibility towards the public at large, whom they sought to protect from harm through recidivism or as a part of the broader criminal justice infrastructure, but felt that this was also harder to discharge given the difficulty of distinguishing "relevant" from "irrelevant" signals in the data. The responsibility that they felt that they could still discharge, and which occupied an ever-greater amount of their time, was one towards organizational processes, or even to the data itself. They suggested that, while the other responsibilities may be beyond their reach, they could ensure that organizational accountabilities (such as the requirement to document and account for problematic data signals) be maintained, even if this necessitated a shift in their sense of their own responsibilities and their professional role.

This case provides some anchoring for a broader inquiry into questions of the relationship between narrative and data. We see multiple actors engaged in narrative acts of different sorts: around the data, with the data, before and after data, in line with or in contradiction to data, and more. Further, these acts of narration tell different stories for different purposes in different moments. In some cases, it is only through narration that the data can speak; in others, narration extends the data's reach. We find, too, people struggling with the problems that what was once narrative is now data, and that the data must be treated or responded to in ways quite different than narrative might once allowed. Finally, we also note the work involved in making data work: as Thornham and Gómez Cruz note elsewhere, "discursive, operational and material constructions of data obscure and mask the enormous effort surrounding data that is necessary to position it as self- legitimating and self-fulfilling" (Thornham and Gómez Cruz, 2016: 8).

This example begins to reveal the way in which sensor-derived data nonetheless needs to be accounted for, both in terms of its production and in terms of its consequences, within social settings. In seeing or framing data as a trace of an event or an action, we inherently invoke narrative elements: actors, motives, expectations, actions, types, histories, proclivities, habits, intents, and on. It is these elements that help us make sense of data as it moves around in the world: as it moves from technical settings into social or organizational ones, for example, or as it moves between different institutions. In these settings, data are narrated differently, and made to operate within different interpretive frames. By the same token, within each of these settings, conventional (if evolving) sets of tropes are invoked in order to make sense of different data streams; we might expect to see different data streams generated by domestic appliances and embedded devices linked by an "Internet of Things" in different homes, but we nonetheless expect some common narratives of domestic life to appear—narratives of rhythm and routine, of intimacy and care, of chores, celebrations, and sleep.

However, the goal of this paper is not to explore a specific ethnographic case of the entwining of data and narrative, but rather to explore more broadly how it is that data practices and narrative practices are entwined, and with what consequences. So we want to move here from this material to a broader conceptualization, albeit one that remains thoroughly grounded in ethnographic methodology.

An ethnographic perspective

There are a number of reasons why, as a group of researchers with a particular interest in ethnographic methods, we might take up this topic—and similarly,

a number of reasons to believe that ethnography has some particular insight to offer here.

One reason is that ethnography proceeds not just by telling stories but by tracing them, analyzing the spread of ideas, expressions, attitudes, and ways of thinking through organizations, communities, and cultures. A sensitivity towards the work that stories do and the way that they are used to make sense of goings-on, to orient others towards shared concerns, and to develop a collective repertoire of thought is at the heart of the ethnographic enterprise, and so it is perhaps not surprising that, as ethnographers, we should be particularly attentive to these elements of data practice. Further, ethnographic approaches provide a set of tools for tracking and unpacking the way that data and narrative are bound together.

A second aspect of ethnographic practice that renders it particularly useful here is its attentiveness to the processes by which inert objects are enlivened, and particular things—objects, places, practices, and ideas—are made to "live" within cultural settings. Proceeding from the recognition that the meaning of cultural things is not fixed but ongoingly produced by people in the ways in which they talk about, make appeal to, explain, contest, celebrate, and debate the significance and values of those things, ethnography sets about unpacking those practices by which meanings are produced at particular times and in particular places. The enlivening of data and data sets as they are mobilized in and through narratives falls squarely within the ethnographer's standard operating procedure. It is not by chance that there is an increasing ethnographic interest in studying algorithms, an essential element on how the narrative about data being "objective" is shaped (Dourish, 2016; Gillespie, 2013; Seaver, 2017).

Finally, here, ethnographers have been particularly sensitive to the ways that their research practices are not just broadly scholarly but particularly literary and textual. The -*graphy* of ethnography renders the writing process indivisible from the research endeavor, and questions of voice, authority, partiality, ownership, perspective, and polyvocality have been widely debated as they arise in the production and dissemination of ethnography's scholarly outputs. So the idea that data must be narrated to be made to work within specific communities and in particular moments is not framed here as a criticism or a problem to be resolved; rather, the very inescapability of these concerns demonstrates how important it is that they be examined and taken up as objects of attention.

Elements of data narrative

A turn to narrative here is essentially a semiotic move. Rather than see data as indexical—that is, as existing by dint of an actual relation to events or objects—we start to see them here as symbolic, and so as taking on meaning through processes of interpretation, "translation" (Bolin and Andersson Schwarz, 2015) and "framing" (Markham, 2013).[3] The process of narration is one by which data is found to be meaningful, and indeed, as Genevieve Bell notes, data has responsibilities: "a story it was compelled to say" (Bell, 2015: 19). Taking a step back, then, we can begin to ask, in what ways do data and narrative interact, and with what consequences? What conceptual commitments come along with a narrative practice in data-centered settings? How does data come to signify, and with what patterns and constraints? These commitments and consequences help to make clear why thinking about data and narrative together provides value within the social study of Big Data and digital data practice.

Data trajectories

One reason to be especially attentive to the relationship between data and narration lies in the way in which narrative begins to add a structuring element and most especially a sequencing element to data. Media scholar Lev Manovich has argued that data-driven media rely upon an associational mode in contrast to the linear mode of the novel and the visual mode of cinema (Manovich, 2001). Narrating data reintroduces a notion of sequence and, with it, a notion of path, of movement, or of trajectory. Trajectory, here, reflects the logic of equivalence, as noted earlier, in the sense of the way in which different elements, data items, or moments are interpreted as being aspects of a singular whole; but of course it also introduces a teleological component, a directedness in which prior actions anticipate future outcomes.

For instance, within the narratives of parolees, the metaphor of the path features strongly. This is not simply the geospatial narrative of movement through different kinds of urban space, but also the broader metaphor of the journey that must be undertaken, the "straight and narrow" path from which the parolee must not deviate, the path of the parolee's progression that leads either upward (with struggle and effort) or downward (in the case of a lack of resistance). Narrative and linearity are deeply entwined and, as Ingold has extensively demonstrated, "the line" is hard to escape (Ingold, 2007, 2016). But too ready a commitment to linearity—conceptual, metaphorical, and spatial—means that data stories can run the risk of being "just so." Linearity belies the complexity of multiple perspectives and alternatives that live within data, and emphasize the selection of particular points of view and the de-emphasis of others. Further, and again drawing on the case of the parolees, we can

recognize too that data stories are not individual, unique, or singular; they embody metaphors and tropes that connect them together. Just as stories are patterned and reflect narrative expectations, embody archetypes, and express conventions, so too do data interpretations. In other words, the character of data narratives here is not merely linear but teleological.

We draw attention to these issues not (purely) to critique the notion of data stories and narrative practices in data science, since, as we have argued, these are inescapable. We turn to them instead in order to observe the way that, once we turn a narrative lens upon data practice, we need to recognize the ways that stories work and what kinds of narrative elements are animated in and around data. Indeed, even at the most reductive level, one might imagine the utility in cataloging data tropes, themes, and paradigms—the signal outlier, the binomial distribution, the emerging cluster, the central tendency, the figure/ground reversal and the broad array of thematic elements that animate TED talks and information visualizations (c.f. Passi and Jackson, 2016). We might see these as aspects of a "professional vision" of data science, to use Goodwin's (1994) terminology, but also as key elements of disciplinary mechanisms of narrative sensemaking (Weick et al., 2005), building on long histories of visual knowledge production (Drucker, 2014; Kennedy et al., 2016).

Drawing attention to the diversity of interpretations in and around data, Fiore-Gartland and Neff (2015) introduce the idea of "data valences," a term that captures "people's expectations of and values for data that emerge from their discourses and practices across different contexts." Drawing on cases particularly in the domain of health care, they detail the entwining of different subject positions and orientations towards data with the particular kinds of values and concerns that data can be seen to support, often yielding contradictory or incommensurate positions on the same data or data set. This is very clear in the anchoring example since the same dataset is useful for both narratives, that of control by the parole office perspective and that of an accountability of freedom from the parolee side. Whether and how data is useful, accurate, actionable, or true is not simply here a matter of perspective; it is more broadly a social relation.

So these are narratives not just *in* data (although they may be) or *from* data but also narratives *of* data. Schrock and Shaffer (2017) build on Fiore-Gartland and Neff's analysis but link it further to Ilana Gershon's (2010) notion of "media ideologies". Gershon's analysis concerns the ways in which people make evaluations of the use of appropriateness of different communicative media for different social interactions. For Schrock and Shaffer, this suggests a parallel notion of "data ideologies": "culturally and socially inscribed beliefs about the appropriateness of data for certain communicative purposes." (p. 2). Their work is grounded especially in examinations of civic data projects and government open data initiatives, in which data is conceptualized as a site for civic engagement, citizen outreach, political communication, and organizational accountability.

Data temporalities

Understanding how both data and narratives are embedded within their own histories might alert us too to other questions of data's temporalities that shape the stories that data tells. Two concern us here: data and narrative dynamics and data-driven futures.

In the realm of Big Data analytics, dynamism is a key component. Data is not simply fixed in place; it is being generated continually and it constantly shifts and develops. So central is this idea to the data rhetoric that "velocity" is one of the four "V"'s by which data science researchers characterize the conditions of "Big Data" analysis (the other three being volume, variety, and veracity). Velocity here speaks to the rate at which data streams are generated and must be processed—the idea that behind each data item is another and yet another, coming quickly. This implies that the rates of data processing must be matched to the rates of data generation, but it implies too that each item is meaningful primarily as an element of a sequence, fast-paced and dynamically evolving.

In this, then, data narratives help to "fix" data temporally. That is, the accounts that data narratives offer are ones that make sense of data within an evolving context, and so stabilize it in the sense that they situate it within a landscape of recognizable objects. This is not to suggest that the data becomes immutable or unchanging, but rather that it is rendered stable and accountable within the terms that a narrative offers; it may evolve and change but it does so within a stable frame. Data narratives help to stabilize data by shifting the temporal scale and giving data meaning even in advance of the inevitable arrival of new, unknown, and unknowable signals.

The last issue of data and temporalities that we address here concerns the way that data is conjured as a means to provide insight into future events, as a matter of both projection and prediction—the idea of data-driven futures.

Of course, an account of data narratives might question the idea that we are data-driven at all, no matter what the dominant rhetoric. Are we driven by the data, or by the stories that the data lets us tell? Are we oriented towards data, or are we oriented towards the narrative logics from which that data springs, and

through which it comes to have relevance? What does an orientation towards data open up and what does it obscure? In the context of an examination of data narratives, the role of data as a "driver" becomes a matter of considerable potential dispute.

However it is that "data-driven" accounts are themselves as much products of narrative processes and logics as they are of data itself, our turn towards topics of "data-driven futures" speaks to the imagined trajectories of data and to the drives for projection and prediction. Data in this account is understood to signal or organize itself into patterns that project into the future as well as the past; data anticipates. However, these anticipations point backwards as much as they do forwards when they are couched or narrated in terms of pre-figured objects and categories.

The cultural grounding of data narratives

A critical point to emphasize in both the anchoring case and others is that the stories that arise in these settings do not rely solely on the data. Elements of these narratives are pre-figured. So it is always with stories; stories operate in terms that we recognize and that are culturally available to us, and so classical narrative forms—of enlightenment and of fall, of struggle and of transcendence, of emergence and of transformation—are broadly culturally available and emerge in conversation with the settings and moments of narrative production. The logic of correspondence, described at the outset of the paper, operates in terms of culturally grounded categories that frame how data anticipates interpretations. For instance, in the anchoring case of parole officers, we noted the ways in which the stories that the data supported were stories that were already available to parole officers about types of offenders, likely patterns of behavior, trajectories of action, traditional forms of danger, sorts of places where people might be, the kinds of risks that they might encounter there, and so on. Each pass through the data tells a story that's new, but the stories are populated and furnished with familiar elements. Similarly, cultural groundings of narrative establish not only conventions of presence, but also conventions of absence—those elements and aspects of the account that are traditionally left out, neglected, or placed to one side. A narrative about movement, for example, is simultaneously and pointedly not a narrative about age, about race, or about gender. What we referred to earlier as the logics of equivalence and correspondence depend thoroughly on these cultural groundings in order to operate; it is only with respect to these groundings that individual data elements can be said to be "about" something, and indeed, "about" the same kinds of some things.

What this further brings to our attention are the histories and geographies of data and those of data narratives. As accounts of phenomena in the world, particular forms of data—such as the reports of latitude and longitude that are the foundation of GPS tracking—are embedded within regimes of measurement and management, with their own histories. Data formats and data representations co-evolve with programs of data use and with anticipated needs. Data has their own histories and their own geographies too, since these regimes of measurement and management are unevenly distributed in the world and are often used to produce logics of spatial experience (the regularization of space through latitude and longitude sits uncomfortably, for example, with indigenous Australian accounts of space that are grounded in relational experience, radiating centers of power, and contemporary encounters with ancestral events—Munn, 1996). At the same time, data narratives, the stories that we can tell about, though, and with data, have their own histories and geographies. They reflect understandings and experiences that have grown up differentially in different parts of the world, or that differently reflect the experiences grounded in gender, sexuality, ethnicity, and economic status, to name a few. Further, these narratives have different ways of moving around in the world, through their proximity to or currency within groups with differing access to media and channels of distribution, from the microwave antenna to the epic poem. Further, and crucially, these two patterns of histories and geographies, those of data on the one hand and data narratives on the other, are not themselves the same.

What this suggests is that the process of holding data, with its histories and geographies, together with data narratives, with its own embeddings, is always both provisional and fraught; a temporary alignment that is always destined to be torn apart as both data and narrative evolve. The stories that data can tell are always stories here-and-now, stories that reflect specific perspectives that may look quite different in the morning.

Conclusions

The argument that data has essentially supplanted theory—explicit in popular writing, but still present, if implicit, in much adoption of data-driven technologies—is the idea that launched a thousand data science programs (c.f. McKie and Ryan, 2015). From predictive policing to diagnostic visual analytics, data is imagined to speak for themself. This is a claim, of course, that has been subject to significant critique from the social science community, with examinations of problems of bias (Angwin et al., 2016), of error

(Garnett, 2016), of locality (Bartlett et al., 2018), of power (Currie et al., 2016), and of opacity (Burrell, 2016).

We share these concerns with the limits of data but seek to bring into focus the pragmatics of data practice. In particular, we have been concerned with the processes of sense-making in and around data, and the logics of equivalence and correspondence upon which data sense-making depends. Data makes sense only to the extent that we have frames for making sense of it, and the difference between a productive data analysis and a random-number generator is a narrative account of the meaningfulness of their outputs. Moreover, one of the most powerful narratives about data is precisely that it demands no interpretation or narration because of its self-evidentiary character. Bringing to this conversation our experience as ethnographic researchers, and as part of a broader investigation of relationships between data and ethnographic methods, we have been especially concerned here with questions of narration.

The particular significance of the narrative perspective is both how it animates a series of culturally-available tropes—actors, motives, encounters, and on—and also how it lends a temporal arc to data and the objects that the data is read to represent. These speak importantly to the cultural embeddings of data narratives, and perhaps to questions of "decolonizing" data (c.f. Smith, 1999) or at least recognizing the importance that these embeddings play in the creation of meaning and the mobilization of action around data that might otherwise seem to speak for itself. These concerns shape not just the encounter with data or with the ways in which data are put to work; they concern too how data is imagined to flow, with how data is seen to represent, and with the ways in which data processing is understood and enacted (Passi and Jackson, 2016).

Indeed, the narrative power of data *in itself* is remarkable. As Loukissas and Pollock (2017) argue, the upset of the 2016 US presidential election, and the difference between the results and pre-election statistical analyses and predictions, did not, it seems, weaken people's confidence in data, despite the fact that the data had let them down. Rather, data's allure remained intact, as commentators argued that, if the data did not anticipate the results, then we must simply not have had enough of it, or not have had the right data. Here, data assumes a role itself in a broader narrative.

The approach we have taken here is one that is particularly ethnographic not in the sense that it has arisen through an ethnographic investigation but rather in that it is informed by an ethnographic outlook. So, in exploring narration in data practices, our goal is not to casually undermine data efforts but rather to resituate them within social circumstances and cultural settings (Neff et al., 2017). A focus on narrative does not necessarily imply falsity; stories, after all, often conjure fictional worlds in order to tell deeper truths. What is interesting instead is how narrative works, and then how it is put to work within the context of data. Taking a cue from Latour (2007), we are interested in the way in which data itself may not sustain truth claims without a narrative framework to make it effective. At the same time, when narrative enframes data, it does so in ways that are pre-figured and embedded in particular locales, cultural settings, and moments in time. As ethnographic researchers, we see much opportunity in the study of narrative opportunities and limits in analysis of data and its social contexts, and indeed, as suggested by Ford (2014), this may itself constitute a new basis for collaborative engagements between ethnographers and (other) data scientists.

Notes

1. These workshops drew on a varied group of participants. Beyond ourselves, the participants at the workshop on data narratives included Heather Horst, Deborah Lupton, Sarah Pink, John Postill, Shanti Sumartojo, and Deb Verhoeven, whose contributions are gratefully acknowledged. To read more about these discussions see https://dataethnographies.com.
2. This research group included Simon Cole, Charlotte Lee, Irina Shklovski, Jennifer Terry, and Emily Troshynski.
3. Arguably, this could be framed as a shift from seeing data as being indexical in a Piercian sense to seeing it as indexical in an ethnomethodological sense (Garfinkel and Sacks, 1970).

References

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias. *ProPublica*. Available at: http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750 (accessed 10 June 2016).

Bartlett A, Lewis J, Reyes-Gallinda L, et al. (2018) The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -omic biology and high-energy physics. *Big Data & Society* 5(1): 1–15.

Bell G (2015) The secret life of big data. In: Boellstorff and Maurer (eds) *Data, now Bigger and Better*. Chicago, IL: Prickly Paradigm Press.

Boellstorff T (2013) Making big data, in theory. *First Monday* 18(10). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750 (accessed 10 June 2016).

Bolin G and Andersson Schwarz J (2015) Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society* 2(2): 2053951715608406.

Burrell J (2016) How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.

Currie M, Paris B, Pasquetto I, et al. (2016) The conundrum of police officer-involved homicides: Counter-data in Los Angeles County. *Big Data & Society* 3(2): 1–14.

Dourish P (2016) Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3(2): 2053951716665128.

Drucker J (2014) *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press.

Fiore-Gartland B and Neff G (2015) Communication, mediation, and the expectations of data: Data valences across health and wellness communities. *International Journal of Communication* 9: 1466–1484.

Ford H (2014) Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data & Society* 1(2): 1–3.

Gabrys J, Pritchard H and Barratt B (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society* 3(2): 1–14.

Garfinkel H and Sacks H (1970) On formal structures of practical action. In: McKinney and Tiryakian (eds) *Theoretical Sociology: Perspectives and Developments*. New York, NY: Appleton-Century-Crofts, pp. 338–366.

Garnett E (2016) Developing a feeling for error: Practices of monitoring and modelling air pollution data. *Big Data & Society* 3(2): 1–12.

Gershon I (2010) Media ideologies: An introduction. *Journal of Linguistic Anthropology* 20(2): 283–293.

Gillespie T (2013) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167–193.

Goodwin C (1994) Professional vision. *American Anthropologist* 96(3): 606–633.

Ingold T (2007) *Lines: A Brief History*. London: Routledge.

Ingold T (2016) *The Life of Lines*. London: Routledge.

Kennedy H, Hill RL, Aiello G, et al. (2016) The work that visualisation conventions do. *Information, Communication & Society* 19(6): 715–735.

Latour B (2007) *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford University Press.

Loukissas Y and Pollock A (2017) After big data failed: The enduring allure of numbers in the wake of the 2016 US election. *Emerging Science, Technology, and Society* 3: 16–20.

McKie L and Ryan L (eds) (2015) *An End to the Crisis of Empirical Sociology?: Trends and Challenges in Social Research*. London: Routledge.

Manovich L (2001) *The Language of New Media*. Cambridge, MA: MIT Press.

Markham AN (2013) Undermining 'data': A critical examination of a core term in scientific inquiry. *First Monday* 18(10). Available at: http://firstmonday.org/article/view/4868/3749nos (accessed 10 June 2016).

Martin A and Lynch M (2009) Counting things and people: The practices and politics of counting. *Social Problems* 56(2): 243–266.

Mayer-Schoenberger V and Cukier K (2013) *Big Data. A Revolution that will Transform How We Live, Work, and Think*. London: John Murray.

Munn ND (1996) Excluded spaces: The figure in the Australian Aboriginal landscape. *Critical Inquiry* 22(3): 446–465.

Nadim T (2016) Blind regards: Troubling data and their sentinels. *Big Data & Society* 3(2): 2053951716666301.

Neff G, Tanweer A, Fiore-Gartland B, et al. (2017) Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data* 5(2): 85–97.

Ortner S (2006) *Anthropology and Social Theory: Culture, Power, and the Acting Subject*. Durham, NC: Duke University Press.

Papacharissi Z (2015) The unbearable lightness of information and the impossible gravitas of knowledge: Big Data and the makings of a digital orality. *Media, Culture & Society* 37(7): 1095–1100.

Passi S and Jackson S (2016) Data vision: Learning to see through algorithmic abstraction. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Media,* San Francisco, CA, February 2016, pp. 2436–2447.

Pink S, Ruckenstein M, Willim R, et al. (2018) Broken data: Conceptualising data in an emerging world. *Big Data & Society* 5(1): 2053951717753228.

Schrock A and Shaffer G (2017) Data ideologies of an interested public: A study of grassroots open data intermediaries. *Big Data & Society* 4(1): 1–10.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 2053951717738104.

Sharon T and Zandbergen D (2017) From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media & Society* 19(11): 1695–1709.

Shklovski I, Troshynski E and Dourish P (2015) Mobile technologies and spatiotemporal configurations of institutional practice. *Journal of the Association for Information Science and Technology* 66(10): 2098–2115.

Shklovski I, Vertesi J, Troshynski E, et al. (2009) The commodification of location: Dynamics of power in location-based systems. *Proceeding of International Conference on Ubiquitous Computing Ubicomp 2009,* Orlando, FL, September 2009, 11–20.

Smith LT (1999) *Decolonizing Methodologies: Research and Indigenous Peoples*. Dunedin, NZ: University of Otaga Press.

Strathern M (2003) Commons and borderlands: Working papers on interdisciplinary, accountability and the flow of knowledge. Sean Kingston.

Thornham H and Gómez Cruz E (2016) Hackathons, data and discourse: Convolutions of the data (logical). *Big Data & Society* 3(2): 1–11.

Troshynski E, Lee C and Dourish P (2008) Accountabilities of presence: Reframing location-based systems. *Proceedings of ACM Conference on Human Factors in Computing Systems CHI 2008,* Florence, Italy, pp. 487–496.

van Dijck J (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

Veel K (2018) Making data sing: The automation of storytelling. *Big Data & Society*. Epub ahead of print 2018. DOI: 10.1177/2053951718756686.

Weick KE, Sutcliffe KM and Obstfeld D (2005) Organizing and the process of sensemaking. *Organization Science* 16(4): 409–421.