Maintaining Data Infrastructures

STEVEN J. JACKSON, Cornell University;
JEN LIU, Cornell University;
RANJIT SINGH, Data & Society Research Institute; and
SAMIR PASSI, KForce.

Suggested citation for this draft book chapter completed on 15 September 2023:

Steven J. Jackson, Jen Liu, Ranjit Singh, and Samir Passi, 'Maintaining Data Infrastructures', in Tommaso Venturini, Amelia Acker, Jean-Christophe Plantin, Antonia Walford (eds.), *The SAGE Handbook of Data and Society: An Interdisciplinary Reader in Critical Data Studies* (Thousand Oaks, CA: Sage Publishing, forthcoming).

Keywords:

data, infrastructure, maintenance, repair, breakdown, sustainability, labour, fairness, inequality.

Abstract:

This chapter explores the central but oft-overlooked role of maintenance and repair in data systems and infrastructures of all sorts. It considers questions of who, what, where, when and how maintenance work is accomplished, and the key part played by maintenance in both establishing and sustaining data projects of all kinds. Drawing on cases from network restoration, soil archives, government databases, and AI models, it considers maintenance work as an underappreciated site of labour, materiality, conflict, and future-making with real and consequential effects on both data infrastructures and the worlds they touch and shape.

Introduction

The work of maintenance has always been central to the practice of infrastructure. In worlds of science, the maintenance of data, equipment, places, people, epistemic authority, and worldviews has begun to achieve the level of attention and respect once reserved for more hallowed terms like Discovery, Theory, and Knowledge. In data and computing, attention to the maintenance of software, hardware, and networks has begun to displace (we hope?) more abstracted and 'cloudy' descriptions of computing as an enterprise. As this new perspective insists, data infrastructures *do not exist* without maintenance: behind every successful system stands a mechanic with a wrench (or data manager with a spreadsheet?). This recognition has made its way into core understandings of infrastructure itself. As Ruhleder and Star classically observed, "*the normally invisible quality of working infrastructure becomes visible when it breaks*" (Ruhleder & Star, 1996, p. 133). This observation underlines a concept central to both American pragmatism (cf. Dewey, 1922) and German phenomenology (Harman, 2002): namely, the world-disclosing properties of breakdown and failure. But it also signals two additional and essential points: the everywhere-and-continuous work of maintenance and repair by which such outcomes are forestalled and systems and worlds are kept going, one fraught fix at a time (Jackson, 2014; Henke & Sims, 2020; Denis & Pontille, 2022); and the world-*making* (not just restoring) properties of maintenance and repair: the fact that maintenance builds worlds forwards as well as backwards, and is thus central to the forms of promise and imagination that infrastructures (including data infrastructures) are so frequently caught up in (Anand et al., 2018; Carse, 2014; Larkin, 2013; Ureta, 2015). Breakdown and maintenance also separates and distinguishes: consider the racially disparate outcomes of automated face or voice recognition systems (Buolamwini & Gebru, 2018; Koenecke et al., 2020), or the treatment of 'low resolution citizens' under universalizing identity schemes like Aadhaar (Singh, 2020b; Singh & Jackson, 2021). The divergent effects of breakdown are therefore central to Star and Bowker's relational reading of infrastructure, with its concern for silences, residuals, invisibilities, and 'orphans' (Bowker & Star, 1999; Star & Bowker, 2007).

But what does it mean to maintain data infrastructures? Who does this work, when and where does it happen, what exactly is being maintained, and how can a better understanding of maintenance change how we imagine, do, and live with data infrastructures today? As used here, data infrastructures refer to all the pieces—physical, virtual and human—that create, operationalize, and preserve data and its specific relations to its world(s) as meaningful, usable, coherent and more or less durable entities, capable of being taken up and deployed in diverse human projects. Foregrounding these relationalities shifts our attention from infrastructures as fixed and accomplished *objects* (infrastructure as noun) to data infrastructur*ing* as a kind of ongoing accomplishment sustained through processes of maintenance and reinvention (infrastructure as verb) (Pipek & Wulf, 2009). This move extends our understanding of data infrastructures in several important ways. It helps us to see maintenance as a kind of thread that connects data infrastructures to the shifting worlds around them, establishing a representational fealty or fit that is *earned*, not given, between data and the phenomena they seek to describe (including through crucial relations of trust or authority that hold data credible and actionable despite their myriad limits, uncertainties, and failures (Passi & Jackson, 2018)). It emphasizes our lived experience of data worlds by moving beyond relations that are purely functional or thinly epistemic to consider investments of care, labor, and value that bind and support long-term data programs of all kinds (for example, in ecology and the earth sciences (see inter alia (Likens, 1989; Lindenmayer & Likens, 2011; Puig de La Bellacasa, 2015; Ribes & Finholt, 2009; Ribes & Jackson, 2013; Baker & Karasti, 2018). It helps us see how data infrastructures manage human actions through time and space, securing the integrity and provenance of data as it moves from past and present to anticipated futures; and from sites of collection (that can also be extractive or colonial in nature) to sites of analysis, application and use (Monteiro, 2023). Through all of this, the 'object' at the center of the enterprise remains subject to subtle and sometimes radical transformation: if data are indeed never raw (Gitelman, 2013) they are also never fully and finally cooked (in the sense of immutable, self-standing, and timeless). These observations should condition and extend our sense of data itself, helping us to understand how data operates within a wider field that, despite enduring tropes of innocence or neutrality, is deeply sociopolitical and sociomaterial in nature, and inevitably caught up in the relations of power and priority that order the accomplished world around us.

This entry explores the maintenance of data infrastructures as a site and practice by which the meaning, integrity, and coherence of data is perpetually achieved. We describe maintenance's complex entanglements with material and earthly processes— all too often erased under the abstractions that limit our discussions of data and computing—and insist on the essential continuity between the 'informational' and the material aspects of data infrastructures. We also show how the maintenance of data infrastructures shapes the emergent conditions of living in a data-driven world, and is essential to the lines of in and out, identity and misrecognition, that shape, define and determine the data worlds around us. Drawing on cases from soil archives to AI models, network recovery to biometric databases, we follow the multiple forms of work and workers by which data is kept live, intact, and valued across space and time; the crucial lines of inclusion and exclusion that constitute the essential margins of data infrastructures; and the fragile forms of order and ordering that the maintenance of data infrastructures enacts and upholds, all against the backdrop of a dynamic and changing world. Vis-a-vis wider discussions in this volume, we point to the error of data as a free-standing noun: there is no data without data infrastructures, and infrastructures are always a verb. Maintenance is what keeps them going, and how they change and evolve into the future.

Exploring maintenance: the who / when / where / what / how of maintenance

As the sensibility above describes, failure, breakdown and decay are the omnipresent condition and natural state of infrastructures writ large, short of a great deal of work to the contrary. Bridges break, foundations erode, pavement cracks, and components age and degrade, eroding the standing and performance of physical infrastructures from roads to rail systems to the buildings that support and constitute urban life (Graham & Thrift, 2007). The same is no less true of data infrastructures, from networks and archives to databases and models— a point widely missed in common cultural stories about data, with their tendencies towards weightlessness, abstraction, and emphasis on new technique or design. The result is a double erasure: of the maintenance work and workers that produce the continuity and endurance of these systems; and of the real risk and precarity that confronts these systems should maintenance be deferred, withdrawn, or chronically underfunded. We open by reviewing answers in the literature to three central questions: *Who does maintenance work? When and where does maintenance happen? And what gets maintained (and how)?*

Who does maintenance work? Maintenance is widespread, and lots of people do it (though rarely in a fully credited or acknowledged way). As a long line of work in feminist STS and data studies has shown, maintenance work all too often constitutes a kind of "invisible labor" or "shadow work" that goes underrecognized, undervalued, and undercompensated (D'ignazio & Klein, 2020; Illich, 1980; Shapin, 1989; Star & Strauss, 1999). These invisibilities frequently coincide with other lines of inequality and exclusion, with essential roles such as cleaners, health care workers, teachers, and more often filled disproportionately by women, migrants, and people of color (Duffy, 2007; Glenn, 1992; Palmer, 2010; Pellow & Park, 2002). The same patterns appear in our data infrastructures: for example, in the gigwork economy, where formerly integrated tasks (including a great deal of basic data cleaning) are decomposed, deskilled, and contracted out via apps and platforms like Amazon Mechanical Turk to a shadow army of piecework employees who lack the rights, powers, and securities of a more formally organized workforce (Irani & Silberman, 2019; Gray & Suri, 2019; Qadri, 2021). At the device level, a small army of repair workers around the world, from authorized service centers to independent shops and markets, keep phones and computers working against the vagaries of hardware and software fails, water, dust, drops, glitches, and a thousand other maladies (Ahmed et al., 2015; Houston, 2019; Jackson & Houston, 2021; Jang et al., 2019; Lu, 2022; Nemer & Chirumamilla, 2019). Data centers are kept running by work and workers (also water and carbon) obscured behind mystifying metaphors like 'the cloud' (Monserrate, 2022; Mosco, 2015). Large-scale science efforts rely on the essential but unheralded work of data managers tasked with cleaning, preserving, updating, and migrating data across the myriad researchers, field sites, and platforms of a distributed and long-term scientific enterprise (Ribes & Finholt, 2009; Karasti et al., 2010; Baker & Karasti, 2018; Thomer & Rayburn, 2023). That these sites are *not* well accounted for in our general stories of data in both academic and popular venues is both an ethical and an epistemic loss: as such activities are rendered invisible, undercounted and undercompensated, the unique insights and understandings available to this positionality—what we have described elsewhere as a standpoint epistemology of repair (Jackson, 2014)—are lost or obscured, often to the detriment of the sustainability and success of the enterprise as a whole.

When and where does maintenance happen? As a form of work deeply grounded in social and material context, maintenance is specific and accountable to both time and place. Some of these activities are reactive, responding to failures and breakdowns suddenly appearing: the code which no longer runs, the screen suddenly cracked or pixellated, the '404' file error indicating a broken link structure. Others are anticipatory, preserving data infrastructures against known or expected breakdowns—for example, efforts to head off widespread software and system failures associated with the Y2K problem, or the efforts now underway throughout the world to 'harden' network infrastructures against the effects of sea level rise, extreme weather events, and other forms of climate disruption (Bijker, 2007). Maintenance is thus both routine and extraordinary—the everyday and 'boring' activities that constitute the regular rhythms of systems and organizations, and the sudden or episodic responses that break and reset routine, sometimes onto new paths and courses. Maintenance is also closely linked to processes of adjustment and adaptation *through time*: the myriad tweaks and fixes by which data infrastructures, from models to operating systems, are kept 'working' within specific and shifting environments and contexts of use—and in the process *re*made, brick by brick (patch by patch?) into updated versions of their former selves. This observation runs against a long-running blind spot in cultural and academic stories around technology, which have tended to privilege invention, innovation and design as the central sites of dynamism in technical systems. Scholars from David Edgerton (2007) to Vinsel, Russell and the wider Maintainers Network (Vinsel & Russell, 2020) have sought to reverse this claim, inverting heroic innovation narratives to (re)tell the history of technology as a story of "how a group of bureaucrats, standards engineers, and introverts made technologies that kind of work most of the time" (Vinsel and Russell, 2014).

What gets maintained (and how)? If maintenance often connotes everyday repetition and routine, the real-world character of this work is diverse and multiple, running the gamut from the mundane, boring, and habitual to the creative, improvisational and inspired. It can include the mechanical and functional but also the aesthetic, as when the management or response to error, breakdown, and imperfection comes to constitute an aesthetic all its own—in kintsugi pottery for example, with its delicate gold threads rejoining the cracks of 'broken' ceramics (Kelly et al., 2021; Rosner, 2018), or in the management of error and 'blue notes' in jazz, from which both continuity and a genuine novelty and difference in sound is produced (Berliner, 2009; Klemp et al., 2008). The same principles carry through to data infrastructures, where breakdown and error may function as central sites of learning and innovation (for better and for worse) in things ranging from machine learning and large language models (Lin & Jackson, 2023), to network outages, to the kinds of automatic flagging in content moderation systems that mistake medical imagery and discussion for pornography. At the same time, the breakdowns which occasion maintenance (or reveal its failure) may index and reflect wider problems of order and inequality—for example, automated street directions which routinely read 'Malcolm X' as 'Malcolm Ten' (Benjamin, 2019) or the problem of residual categories described by Bowker and Star (1999), like the actually existing people rendered legally impossible under the databases and classification schemes of apartheid. These examples remind us that if maintaining data infrastructures is often a sectoral or technical question, it also references some of the widest and most important collective questions we face. What we allow to break and what we choose to maintain is an index of value, priority and care.

As the cursory review above makes clear, the maintenance of data infrastructures is ubiquitous, essential, and performed by a wide cast of actors, not all of whom receive appropriate credit and recognition for their work (and are frequently excluded from narrower and more abstracted notions of data and data science). It is also caught up in and essential to larger projects: both those attached to the immediate purposes of the infrastructure at hand, but also wider projects like the authority and efficacy of the state; the promises of modernity, both realized and unrealized; and the justice and equity (or not) of our basic systems and infrastructures. The cases that follow explore these themes in greater detail, showing how the who/what/where/when/how of maintenance might support a different and wider understanding of data infrastructures and the relations they hold together. Our first case explores the physical and network underpinnings of data infrastructures, and the coordinated efforts to keep data flowing across landscapes that are vulnerable to change. Our second case explores the irreducible materiality of data, and the efforts to preserve and maintain physical archives as a basis and ground zero for long-term (but always evolving) scientific collection programs. Our third case considers the problem of state records and record keeping, and the various forms of work required to maintain and assert individual and categorical identities—including the crucial category and entitlements of citizenship—within data systems that are also prone to glitch, failure and exclusion. Our fourth case involves the maintenance of AI models, and the ongoing effort to anticipate and adjust to model failures before they happen.

Maintaining Grids and Networks (or, How To Restore the Internet After a Storm)

After Hurricane Ida made landfall on August 29, 2021, more than one million people lost power across southeast Louisiana. The storm brought winds recorded at 150 miles per hour, leaving tens of thousands of utility poles knocked over or damaged. Many of these poles were joint poles, carrying both electricity and telecommunications services to residents. At the time, the COVID-19 pandemic also meant that many people were relying on the Internet for remote work, education, and other services that had shifted to virtual environments. Where cell towers remained operational, the backup generators keeping them powered soon ran out of fuel. With roads yet to be cleared, the cell towers were not accessible for service crews to refuel them. The power and Internet outages also made it difficult for people to assess and record damage after the storm to file for federal assistance and insurance claims. For insurance companies and agencies like the Federal Emergency Management Agency (FEMA), data, in the form of documentation and photos, need to be uploaded to their websites. Without a functioning local network, this essential data work was impossible to complete.

One electric utility reported that over 4000 of their utility poles were damaged and needed to be replaced. Most of their networks were still located above ground with lines aerially tethered to utility poles. The utility uses this method (instead of burying lines) partially due to the high cost of burying cable, especially in areas so close to sea level, and the potential to better identify any issues across their service territory. While these poles are shared between telecommunications and electrical infrastructures, electrical utilities often own the poles and are thus tasked with pole maintenance. Electricity lines are placed on the top, as a safety measure, while underneath are the data lines for services like DSL, cable, phone, and fiber optics. This shared placement means that when a storm knocks down a pole, both infrastructures are compromised.

This vertical ordering of wiring meant that the work of replacing poles and restringing lines became a tightly coordinated effort where work must occur in a specific order during the recovery period after Hurricane Ida. Even before the storm made landfall, hundreds of subcontractors and crews from across the state and country were on standby at the request of the utility. Once the storm passed, they made their way to assist with recovery. Debris was first cleared off of roadways. Then, poles were re-installed. When the poles in a section were finally installed, the lineworkers came through to mount the electrical lines to the poles - a choreographed procedure to ensure proper tension across the points of contact. Only afterwards are telecom workers able to come through to reinstall communication lines.

However, repairs can be difficult in certain parts of the service territory. On right-of-way terrain that is more terraqueous, such as swamps, specialized vehicles are needed to help install poles. Workers will often work from the boats or in chest-high muck to place poles into the earth. As a result, these repairs can often be time-consuming, labor-intensive, and expensive. When asked about the decision to place utility poles through a waterway, a line worker responded: “Well it wasn’t always water.” This observation reflects the growing ecological issues that south Louisiana is facing as a region built on primarily fragile wetlands. One of these issues is rapid land loss, where the rate of erosion is outpacing the rate of land renewal. Hurricanes contribute to land loss by causing storm surges that can deteriorate vulnerable land, in addition to inviting in seawater that can disrupt the ecosystem. With stronger storms happening more often due to climate change, getting networks back online requires adapting to a changing landscape.

Networks will inevitably face damage after a storm. However, these circumstances become more dire as the consequences of climate change unfold in south Louisiana. Poles that once stood on solid ground are now in swamps as the water moves closer, swallowing the land bit by bit. Viewing these infrastructures as standalone configurations independent of the places they are embedded in belies the interdependence of infrastructures and the coordinated work required to bring them (back) online. In the recovery work after Ida, repairing transportation and energy infrastructures were crucial steps in restoring the Internet. Pathways needed to be cleared and poles needed to be installed by workers in muddy, humid, and often dangerous conditions. Maintaining data requires holding onto an increasingly unstable ground.

Bucket trucks preparing for Hurricane Ian, Sumter County, Florida, October 2022

Maintaining Material Archives (or, How To Read the Planet Through Soil)

Our second case concerns the materiality of data—and the inevitable and often invisible forms of maintenance work required to sustain it. Consider for example the problem of soil. The Australian National Soil Archive, housed at the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Canberra, contains more than 70,000 soil samples collected across the varied soilscapes of Australia, from farmland to coastal zones to the arid interior. The oldest samples, dating to 1924, preserve a memory of a world before industrial agriculture, synthetic fertilizers, and atomic bombs. They have been used to investigate soil carbon stocks across Australia, soil property change over time, and more specific questions ranging from carbon sequestration to soil salinity, composition, and diversity to forensic investigations by the Australian police. In a manner consistent with other long-term collection programs, many contemporary analyses making use of the CSIRO collections employ techniques (infrared scanning, new modes of carbon measurement) and address concerns (the effects of nuclear fallout; the impacts of climate and atmospheric change) that would have been unimaginable or uninteresting at the time of early collection. The unknown and open-ended future of social and scientific interest remains a central and orienting commitment of the center, and materials are shared out sparingly to conserve this capacity; as CSIRO's own materials note, "besides being an unrivaled resource for addressing many established issues, soil archives have the potential to provide answers to agricultural and environmental questions not yet answered" (Karssies et al., 2011, p. 7).

Seemingly inert and notably light on care and feeding, soil samples would seem to be the Easy Case of maintaining data infrastructures in their physical form. But this turns out to be far from true. The archive contains more than 55000 samples stored in carefully labeled and ordered one liter plastic containers, air-dried or oven-drived to 40 degrees Celsius and drawn from more than 200 different research projects and collection programs around the country, each of which must be calibrated for consistency. (A separate pink card collection includes older soil samples currently listed as 'unarchivable'). Both temperature and relative humidity are carefully tracked and managed. The conditions behind each sample are also carefully recorded, including key data such as location coordinates, soil depth, and extraction method. Beyond the soil, copious amounts of paper are also stored at the archive (some of it in slow and laborious process of being digitized), including the crucial field sheets and collector notes that accompany (some of) the accessed collections. All of this represents just one small example of the vast (!) scientific back room, from assays to tree rings, specimens to ice cores, fossils to sediments, on which knowledge in ecology and the earth sciences—the 'archive' of climate science, as Shannon Mattern (Mattern, 2017) has observed—is ultimately grounded. And because soil collected and preserved to these exacting standards is a rare and finite resource, the entire collection must be managed parsimoniously, ensuring that sufficient quantities are kept against the as-yet unknown needs and interests of the future. Relations with landowners, politicians, and 'the public' too must be maintained, so that the resources needed to support the work continue to flow uninterrupted and the entire operation does not become the victim of budget hawks, science skeptics, or the allure of flashier new science investments.

The samples collected also subtend a rich and complex ecology of data in digitized form, including various standards and measures (of salinity, structure, and chemical composition) and the crucial forms of metadata (entered painstakingly into the 'NatSoil' database) that sustain their meaning and value. This too must be maintained against the backdrop of a dynamic world, as programs move through various standards, file formats, addressing systems, and the like. The building itself must be made resilient to decay, proofed against rot, mold, boring insects, water infiltration, shifting foundations, unwanted plant and animal visitors, and the myriad other forces that afflict and transform modern buildings. And this entire program of work is sustained and reproduced across place, time, and the myriad and changeable humans who pass through it—PIs, government officials, graduate and undergraduate researchers, field techs, etc. (Karssies et al., 2011).

Now consider a strange thought experiment. A vandal enters the CSIRO at night determined to upset this intricate world. (A member of the Soil Liberation Front? An oil operative seeking freedom to drill? A conspiracist from the bowels of the web? The motives are obscure). They pull open the sample drawers, unscrew the lids, and begin pouring out the contents onto the floor, moving up and down the cavernous rows. Hours later, their dastardly work accomplished, they exit the same way they have entered, leaving behind what looks to all the world like a misplaced beach. Moment by moment, a miraculous and terrible transformation has occurred: the contents leave the containers as data, but hit the floor as dirt.

This example (and the strange fantasy which concludes it) suggests other necessary languages in our discussions of data infrastructures: not the fetishization of data so often encountered—big! potent! inexorable!—but also *weak*, *poor*, *fragile* and in some ways remarkably ill-adapted to the worlds it lives in and encounters. It reminds us that data is held in place *as data* only by dint of a great deal of human and extra-human work—by scientists, technicians, carefully controlled storage mechanisms, and hopefully in future better security systems. Data lives within worlds made and kept safe for data. Outside of these worlds, it won't last five minutes—or indeed the time it takes for soil to fall from container to floor.

Maintaining Government Databases (or, How to Claim Citizenship in a Datafied State)

Our third case concerns the everyday work of maintaining the relationship between citizens and their data. Data enables state bureaucracies to represent citizens in organizing their services and citizens to represent themselves to these institutions (Redden, 2018; Singh & Jackson, 2021). Unlike the soil samples, citizens act and complain. Taking action to maintain their own data in government databases is a site of ongoing negotiation with the state with emergent consequences for the life chances of citizens.

Consider Aadhaar, India's national biometrics-based identity database, which was implemented on the promise of providing a unique digital identity for every Indian resident including those who previously lacked identity documents. The bureaucratic work of incorporating Aadhaar into state services was meant to transform ongoing processes of digitizing paper-based citizen records in India (UIDAI, strategy document). Since paper-based records are prone to fraud and forgery, the challenges of uniquely identifying citizens using paper-based documents persist in digital records. Seen in this context, adding an Aadhaar number in an individual's data records in public and private services is an intervention in maintaining data. As an interlocutor explained: "say we are trying to computationally figure out who owns what land in a municipality for taxation. [...] It is impossible to uniquely identify a property owner simply based on their name, so it is also difficult to figure out how many properties a single person owns in a particular municipality, let alone a state or a country like India" (Conversation with a technologist involved in designing e-Governance portal, 29 September 2015 (Singh, 2020b, pp. 121–122)). In computational terms, this difficulty is articulated as the problem of *entity resolution*: the task of matching diverse data records to real world entities, including resolving the problem of different names for the same person in different data records. This task at its core is about maintaining the relation of data to persons or things in the world.

Aadhaar is a macro-scale intervention in maintaining citizen data, but it shifts the responsibility of maintaining the relationship between data and citizens. Entity resolution becomes a matter of individuals (here, citizens) sustaining relations with their own data with the support of street-level bureaucrats. It happens not only when Aadhaar numbers are added to other databases, but also when they are created and used to claim identity. While this work of maintaining data cannot even begin without creating an Aadhaar number, claiming identity through Aadhaar involves navigating diverse challenges ranging from lack of distinct biometric features (for example, the worn, fading or missing fingerprints of manual laborers and the elderly) to infrastructural barriers such as lack of electricity and internet. This story provides a poignant example:

> The tribal areas of Rajasthan [a state in the west of India] are very hilly. How would you get network [internet] there? Let me give you an example. The Fair Price Shop is in the middle of one of these villages in the hills. The ration [subsidized food grains] is there; the shop owner is there; the POS machine [Point-of-Sale machine equipped with fingerprint readers] is there; and so, the villagers went there. When

> the villagers put their thumbs on this machine, it did not work. There was no network. The shop owner decided that the network problems were happening because they are in a valley, so he decided to hike up a hill to look for network. The villagers followed him around with their ration cards to find network on top of the hill. Once they found network, they put their thumbs on the machine again. [...] The elderly and the disabled, of course, cannot walk up a hill. They remain excluded. By chance, if they are successful with [Aadhaar] authentication, they go down to the shop later and collect their ration (A Right to Food activist narrating authentication troubles in accessing subsidized food grains, 26 July 2016, (Singh, 2020b, pp. 274–275)).

While this specific example is from India, the work of claiming identity to avail oneself of state services and entitlements is common around the world. It is not just about proving an existing relationship with a data record in a government database; it is also about *maintaining* that relationship. Entity resolution is a moving target. A pattern of failed efforts to claim a data record may be used as justification to treat the record as a ghost entry and erase it; dormant or unexercised identities may lead to a presumption of death and elimination (Singh, 2020a). These processes, experienced unevenly, can be complex and costly to reverse. Whether it is standing in line at a bureaucratic office to receive a state service or applying to renew a state-issued ID document after its expiration, this work is ordinary and mundane. Yet it is unequally distributed across intersections of gender, race, class, caste, and ability and has profound consequences for citizenship in a datafied state.

Maintaining Models (or, How to Anticipate Algorithmic Failures in the Wild)

Our fourth case involves the maintenance of Artificial Intelligence (AI) models, and the consequential forms of *anticipation work* needed to maintain alignments between data, models, expectations, and wider worlds. AI models have become central components of several data infrastructures and are used to discern relations between people, objects, and practices, learning from the past to provide 'useful' insights in the present. As a growing body of work in human-centered data science has begun to attest (Muller et al., 2019; Passi & Sengers, 2020; Tanweer et al., 2022), it takes a lot of work to build *and maintain* AI models: the AI practitioners' job does not simply end with model deployment (Schmitz, n.d.; Verre, 2019, 2020). Such models are built using historical data (static snapshots of the world). Just like citizen data, while these models become a part of people's lives and practices, the world continues to evolve and grow, at times in uncertain ways; a model that works today may not work tomorrow. For instance, Zillow built its Zestimate® model to predict home prices but the Covid-19 pandemic affected the housing market, causing the model to provide misleading predictions that had drastic repercussions ranging from driving the housing market in an unstable direction to large-scale layoffs (Bahney, 2021; Fu et al., 2022; Pedram, 2021).

A good model catches the world's many drifts (Chiang and Yin, 2021; Lu et al., 2018). Data drifts and so do concepts (e.g., live data starts deviating from training data or the thing you want to predict changes). Maintaining models is thus a constant struggle to align models' view of the world-as-data with the rich tapestry of people, their practices, and other systems. Practitioners spend substantial time and resources on model upkeep—a practice that encompasses work across computational (code, data, models), material (servers, data farms, GPUs), social (user training), and service (APIs and documentation) aspects. From resolving bugs and fixing vulnerabilities to ensuring compatibility and mitigating harms, AI practitioners do various kinds of maintenance to fix issues post deployment. Model maintenance work, however, remains underexplored in current research on data infrastructures—partly because such work is often left out of common portrayals of data science work in academic and research settings

(Passi and Jackson, 2017). Below we take a brief look at model maintenance, focusing on the practice of telemetry to highlight the anticipation work needed to maintain models.
Once a model issue is identified, practitioners fix it using different strategies such as retraining models or blocking the generation of specific outputs. Addressing known failures, however, is but one form of maintenance, AI practitioners must also reckon with how models may fail in subtle, unpredictable, and unique ways, requiring constant monitoring in the wild. One response to this problem is the method of *telemetry*—the proactive collection and analysis of data at points of technology use (e.g., service timeouts or error rate). For AI, telemetry commonly involves measures or variables such as model latency, confidence scores, input distributions, or how often users accept and reject model recommendations (Dhinakaran, 2022; Oladele, 2022). This data is used for various purposes such as monitoring model performance (e.g., precision/recall) and providing performance guarantees (e.g., latency).

Telemetry is a difficult and costly endeavor. Even beyond its technical complexity, designing telemetry requires AI practitioners to take on forms of "broken-world thinking" (Jackson, 2014), recognizing that all models will eventually fail, and preparing to catch the fallout. Then comes the "anticipation work" (Clarke, 2015)—imagining the many ways in which models *can* and *will* degrade, fail, and cause harm. As a form of anticipation, telemetry embodies a core normative ideal of AI practitioners (e.g., what *should* we measure, how *should* models work, what forms of failure are *our* responsibility?). As such, good telemetry is increasingly seen as a hallmark of responsible AI—practitioners keeping a watchful eye on their models, and changing them (perhaps even shutting them down!) if they see them doing more harm than good (Lewis, 2022). The work of anticipating broken and failed futures is an intrinsic—and often emotionally challenging and taxing—part of model maintenance.

The work of anticipating, identifying, and measuring failures and harm remains one of the most challenging aspects of responsible AI practice (Amershi, 2020; Madaio et al., 2022). Telemetry helps, but also brings its own challenges. Take the problem of overreliance on AI—"when users start accepting incorrect AI outputs" (Passi and Vorvoreanu, 2022) —that causes significant harm given AI's use in domains such as software development, medicine, and law enforcement. AI practitioners want users to not blindly rely on models, but instead to develop appropriate trust in them so that they can spot and address model mistakes. But how do practitioners know if and when users over-rely on models? They use telemetry to measure overreliance—e.g., how often users accept incorrect recommendations (Buçinca et al., 2021). However, this is not easy to measure in situations where ground truth isn't available (e.g., when making future predictions). In that case, practitioners rely on proxies such as measuring how quickly users accept recommendations as a potential signal for overreliance. This task again is made difficult when models make different kinds of recommendations. For example, in healthcare, models might predict disease type but also recommend treatment plans—overreliance on one is not the same as overreliance on the other.

Maintenance-via-telemetry occurs at a *distance*; it provides partial windows on AI performance detached from immediate contexts of use. It also operates at *scale,* generating unwieldy amounts of manually intractable data that require practitioners to find effective, often creative, strategies to render it useful. Collecting more data to address 'distance' issues often exacerbates 'scale' issues. Maintenance is thus as much a proactive and speculative practice as it is a reactionary and everyday endeavor. We maintain things in the present but doing so requires understanding how future changes in data and practices may disrupt the delicate temporal relations or "rhythms" (Jackson et al., 2011) that connect data infrastructures with the worlds they mediate. As a form of anticipation, maintenance keeps the world going—often by conjuring different worlds into being, and sometimes by ensuring certain worlds never come to be. Anticipatory practices such as telemetry make visible not only *what* is cared for but also what gets left out and why. Practitioners may (for the right reasons) want to get ahead of problems and proactively monitor model issues, but must do so within the confines of data protection and privacy regulations; the boundary between monitoring and surveillance is slippery. In modeling practice as elsewhere, to maintain is to anticipate failure, degradation, and harm—while simultaneously working to design better ways to capture the uncertainties and fears inherent in the design and use of today's data infrastructures.

Discussion

But what does this strange collection of utility poles, dirt, patchy valley networks, and uncertain model failures, tell us about the wider problems of data and society that frame this volume? Our first point concerns the *sheer immensity and range of labor* involved in maintaining data infrastructures — omnipresent but often invisible (at least to some), in any system that continues to function, data or otherwise. For example, the work of the Louisiana pole crews before and after storms is what keeps the lights on and the data flowing, and makes this particular stretch of coastal Louisiana livable (or livable enough, at least for now). The diversity of our stories should cause us to re-evaluate the basic nature and range of 'data science' work, and the basic metaphors we use to describe and teach it to newcomers. If there's a good deal of 'carpentry' in data science systems (Mimno, 2014)), there's also a decent amount of plumbing, some periodic attention to power outages, and a great deal of work at simply keeping the floors clean. If data infrastructures are to work well and fairly, this labor and skill needs to be acknowledged, valued and transmitted, including through our basic systems of training and education.

Our second point concerns the frequently and irreducibly *material* character of data infrastructure work - and the impossibility of a neat definitional separation between the physical and virtual dimensions of data infrastructures. The case of the Australian soil archive is not a two-world story of (physical) soil and (virtualized) data, but rather a continuous and carefully preserved *line of relation* that links the two—at least until the vandal arrives. Similarly, the story of Aadhaar registration is not only about biometrics and unique data records, but also about signal-blocking hills and legs too old or tired to climb them; the entitlements (or not) of citizenship are what comes out at the *end* of this intricate material-semiotic chain, not the beginning. Data infrastructures are for this reason highly and irreducibly *situated*, arising in concrete places, times, and circumstances (in the rich and literal sense which Michel Serres (Serres, 2008) applies to that term—that which stands around, supporting, upholding). Actors in our stories are constrained and enabled by the worlds around them: hurricanes that push back, signals that drop, housing models too dumb for COVID-19, and outside forces and vandals bent on disorder. And they build on and inherit a prior world of maintenance, with those before them engaged in the same kind of ongoing dance. For this reason, data infrastructures are *always* material, layered, and historical (which is to say, always a work in progress). Maintenance, like turtles, goes all the way down.

Our third point concerns the potentially *agonistic* dimensions of maintenance work—an observation meant to counter a fallacy of neutralism that can otherwise creep into our understanding of data infrastructures and data science as a whole. This point turns on the observation that the forces which challenge, undermine, and erode data infrastructures are not only accidental or environmental, but also sometimes sites of opposition or enforcement built around different/competing notions of what the world could/should mean and be. We have seen above how access to key resources and categories—like the effective exercise of citizenship in the Aadhaar case, or adequate representation in consequential AI models—can turn on problems of breakdown and maintenance. But sometimes systems break or are broken on purpose, expressing a politics just as consequential as those that go into their design and formation. Take as an example the work and complicity involved in maintaining the passbook system (a vast and expansive data infrastructure!) under South African apartheid—and the eventual success of the Polaroid Revolutionary Workers Movement in convincing the company to stop supplying the South

African government with the passbook technology. Or the widespread efforts to defund and undermine climate science, including by rendering public environmental data effectively inaccessible—a move countered by the maintenance work of groups like the Environmental Data Governance Initiative, working to put public data back online. As these examples remind, the maintenance of data infrastructures, like other forms of care work, operates within a fraught and complex political field that can take on and express values and valences of all kinds, some of which we might wish to avoid.

Our fourth point is around the temporal complexity, including the irreducible *futurity*, of data infrastructures. As our stories have shown, maintaining data infrastructures is not just a rearguard action against change, or an effort to hold back the tides of history, time, or decay. It is also an active process of building continuities and bridges to the future, including as we have seen in speculative and anticipatory ways. Thus, the maintenance of network infrastructures in coastal Louisiana is also about maintaining modes of life and livability increasingly challenged under the regional effects of climate change. The technicians at the Australian soil archive must preserve and prepare it against questions, needs and techniques as-yet unknown. When we maintain data infrastructures, we are holding the possibility for (and guessing at?) worlds to come. Put differently: maintenance *builds worlds forwards*, not just backwards (in the sense of an eternal return to a past or origin). The uncertainty of this process makes maintenance work fundamentally creative and *improvisational* in character—and in its widest manifestations, consistent with traditions of endurance, improvisation and hope that can be found in work from feminist and queer theory to Indigenous and Black radical traditions (Jackson, 2023; Moten, 2003; Munoz, 2009; Murphy, 2017).

Finally, and most generally: *maintaining data infrastructures is also about maintaining and (re)shaping the worlds that data touch*—moving us at last, we hope, beyond the myths and misdirections of representationalism that have long challenged work and thinking in data science. Under the relational view advocated here, the relationship of data infrastructures to worlds is not just mimetic or representational, an arms-length reflection or description of phenomena external to and independent of the act of representation. Rather it is *constitutive*, and under the right (or wrong) circumstances, can have a profound and shaping force 'back' on the thing itself—including to the level of its basic constitution and even continued existence. Whether a place is registered as wetland or wasteland matters a lot, including to the ongoing shape and survival of the place itself. How human identities and categories are registered, from the self-asserted to the bureaucratically or diagnostically assigned, can have a profound impact on the character and wellbeing of individual selves and lives. Bit by bit and over time, people and things can become the categories that data infrastructures have devised for them—an effect mostly hidden by seemingly neutral terms like scaling, abstraction, and other mechanisms by which the "blooming, buzzing confusion" of the world (James, 1950) is reduced, channeled and entrained. If data infrastructures "make up people," in Hacking's (1985) apt phrase, maintaining data infrastructures holds them in place, for better and for worse. It can also entrench and deepen the forms of erasure and loss that data infrastructures, like other representational systems, inevitably encode. The outcomes of this process are complex and uncertain: to 'disappear' in data can be a precursor to marginalization, erasure, and existential disappearance in the world. Beyond the bare fact of existence, when things become their categories (and only their categories), a certain kind of loss or "muting" (Ghosh, 2021) has occurred. But to disappear in data, or to exit or refuse the systems that gather and exploit it, can also be a means of preserving freedom, autonomy, the sanctity and distinctiveness of local worlds, and a certain kind of discretionary power, especially in light of the often reductive and 'brittle' character of technical systems (Ackerman, 2000; Star & Bowker, 2007). This is the deep and perpetually unsettled politics of maintaining data infrastructures today.

Author Bios:

Steven J. Jackson is a Professor of Information Science and Science and Technology Studies at Cornell University. His work combines ethnographic, legal, humanistic, and interpretive traditions grounded in pragmatism, critical theory, and post-structuralism with an overall interest in how people build and maintain order, value and meaning in and with the worlds around them. He's especially interested in places where new computing forms and practices meet the social and material worlds, with implications for collaboration, sustainability, justice, learning and inequality. With fellow travelers in the Computing On Earth Lab and around the world, he has written extensively on problems of infrastructure, collaboration, maintenance, repair, and hope.

Jen Liu is a PhD candidate in Information Science at Cornell University studying the ecological, social, and political implications of computing technologies and infrastructures. Drawing from human-computer interaction, science and technology studies, and environmental studies, her research examines how visions and practices of computing are shaped to address pressing environmental and climate issues. She uses qualitative and design methods to understand how computing practices can be redirected towards more sustainable and equitable outcomes. Her most recent work is on the impact of climate change on networked infrastructures in coastal landscapes.

Ranjit Singh is a senior researcher at Data & Society, focused on qualitative research for the Algorithmic Impact Methods Lab (AIMLab) and a research fellow with Siegel Family Endowment. His research sits at the intersection of data infrastructures, majority world scholarship, and public policy for algorithmic systems, and uses methods of interview-based qualitative sociology and multi-sited ethnography. He broadly examines the everyday experiences of people subject to data-driven practices and follows the mutual shaping of their lives and their data records, aiming to understand how data is increasingly used to imagine and develop new digital solutions for democratizing inclusion.

Samir Passi is a social scientist with interdisciplinary expertise in information science, artificial intelligence, and science and technology studies. He is a researcher at KForce, and as part of that position works as a Researcher at Microsoft in the AI Ethics and Effects in Engineering and Research (AETHER) team. He works closely with AI practitioners to address the sociotechnical and organizational issues in the development and use of AI systems. His current research is mostly in the area of Responsible AI, focusing on themes such as overreliance on AI, human-AI complementarity, and red teaming of generative AI-based systems.

References

Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction*, *15*(2–3), 179–203.

Ahmed, S. I., Jackson, S. J., and Rifat, M. R. (2015). Learning to fix: Knowledge, collaboration and mobile phone repair in Dhaka, Bangladesh. *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, 1–10.

Amershi, S. (2020). Toward Responsible AI by Planning to Fail. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3607–3607. https://doi.org/10.1145/3394486.3409557

Anand, N., Gupta, A. and Appel, H., eds. (2018). *The Promise of Infrastructure.* Duke University Press.

Bahney, A. (2021). Zillow to exit its home buying business, cut 25% of staff. *CNN.* https://www.cnn.com/2021/11/02/homes/zillow-exit-ibuying-home-business

Baker, K., and Karasti, H. (2018). "Data care and its politics: designing for local collective data management as a neglected thing." Proceedings of the 15th Participatory Design Conference.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new jim code. *Social Forces.*

Berliner, P. F. (2009). *Thinking in jazz: The infinite art of improvisation.* University of Chicago Press.

Bijker, W. E. (2007). Dikes and dams, thick with politics. *Isis*, *98*(1), 109–123.

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., and Bao, M. (2022). The Values Encoded in Machine Learning Research. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–184. https://doi.org/10.1145/3531146.3533083

Bowker, G., and Star, S. L. (1999). Sorting things out. *Classification and Its Consequences.* MIT Press.

Buçinca, Z., Malaya, M. B., and Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability and Transparency*, 77–91.

Carse, A. (2014). *Beyond the Big Ditch: Politics, Ecology, and Infrastructure at the Panama Canal.* MIT Press.

Chiang, C.-W., and Yin, M. (2021). You'd Better Stop! Understanding Human Reliance on Machine Learning Models under Covariate Shift. *13th ACM Web Science Conference 2021*, 120–129. https://doi.org/10.1145/3447535.3462487

Clarke, A. E. (2015). Anticipation Work: Abduction, Simplification, Hope. In G. C. Bowker, S. Timmermans, A. Clarke, & E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star* (pp. 85–120). The MIT Press.

David Mimno. (2014, August 19). *Data carpentry.* http://www.mimno.org/articles/carpentry/

Denis, J., and Pontille, D. (2022). *Le Soin des Choses: Politiques de la maintenance.* Decouverte: Paris.

Dewey, J. (1922). *Human nature and conduct: An introduction to social psychology.* New York: Henry Holt and Company.

Dhinakaran, A. (2022). *ML Troubleshooting is Too Hard Today (But It Doesn't Have to be That Way).* Towards Data Science. https://towardsdatascience.com/ml-troubleshooting-is-too-hard-today-but-it-doesnt-have-to-be-that-way-fbf9a10cac05

D'ignazio, C., and Klein, L. F. (2020). *Data feminism.* MIT press.

Duffy, M. (2007). Doing the dirty work: Gender, race, and reproductive labor in historical perspective. *Gender & Society*, *21*(3), 313–336.

Edgerton, D. (2007). *The Shock of the Old: Technology and Global History Since 1900.* Oxford University Press.

Fu, R., Jin, G. Z., and Liu, M. (2022). *Human-Algorithm Interactions: Evidence from Zillow.com* (No. w29880; p. w29880). National Bureau of Economic Research. https://doi.org/10.3386/w29880

Ghosh, A. (2021). *The Nutmeg's Curse: Parables for a Planet in Crisis.* University of Chicago Press.

Gitelman, L. (2013). *Raw data is an oxymoron.* MIT press.

Glenn, E. N. (1992). From servitude to service work: Historical continuities in the racial division of paid reproductive labor. *Signs: Journal of Women in Culture and Society*, *18*(1), 1–43.
Graham, S., and Thrift, N. (2007). "Out of Order: Understanding Repair and Maintenance." *Theory, Culture and Society* 24:3, pp 1-25.
Gray, M. L., and Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.
Hacking, I. (1985) `Making Up People', in T.L. Heller, M. Sosna and D.E. Wellbery (eds) *Reconstructing Individualism*. Stanford, CA: Stanford University Press.
Harman, G. (2002). *Tool-Being: Heidegger and the Metaphysics of Objects*. Open Court Publishing.
Henke, C. R., and Sims, B. (2020). *Repairing Infrastructures: the maintenance of materiality and power.* MIT Press.
Houston, L. (2019). Mobile phone repair knowledge in downtown Kampala: Local and trans-local circulations. In *Repair work ethnographies* (pp. 129–160). Springer.
Illich, I. (1980). Shadow-Works. *Philosophica*, *26*.
Irani, L. C., & Silberman, M. S. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–620.
Jackson, S. J. (2014). Rethinking Repair. In T. L. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality and Society* (pp. 221–240). MIT Press.
Jackson, S. J. (2023). Ordinary Hope. In M. Taccheti, D. Papadopoulos, & M. P. de la Bellacasa (Eds.), *Ecological Reparation: Repair, Remediation and Resurgence in Social and Environmental Conflict*. Bristol University Press.
Jackson, S.J., and Houston, L. (2021). Poetics and Political Economy of Repair. in Wasko, J. & Swartz, J., eds. *What Is Media?*. University of Chicago Press / Intellect Books.
Jackson, S. J., Ribes, D., Buyuktur, A., and Bowker, G. C. (2011). Collaborative rhythm: Temporal dissonance and alignment in collaborative scientific work. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work - CSCW '11*, 245. https://doi.org/10.1145/1958824.1958861
James, W. (1950). *The Principles of Psychology* (Vol. 1). New York: Henry Holt and Co.
Jang, E. H. B., Garrison, P., Vistal, R. V., Cunanan, M. T. D., Perez, M. T., Martinez, P., Johnson, M. W., Evangelista, J. A., Ahmed, S. I., and Dionisio, J. (2019). Trust and technology repair infrastructures in the remote rural Philippines: Navigating urban-rural seams. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–25.
Karasti, H., Baker, K. S., and Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW)*, *19*(3), 377–415.
Karssies, L., Jacquier, D., Wilson, P., and Ringrose-Voase, A. (2011). *CSIRO National Soil Archive Manual*.
Kelly, C., Kemper, J., and Rutten, E., eds. (2021). *Imperfections: Studies in Mistakes, Flaws, and Failures.* Bloomsbury Academic.
Klemp, N., McDermott, R., Raley, J., Thibeault, M., Powell, K., and Levitin, D. J. (2008). Plans, takes, and mis-takes. *Outlines. Critical Practice Studies*, *10*(1), 4–21.
Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., and Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, *117*(14), 7684–7689.
Larkin, B. (2013). "The Politics and Poetics of Infrastructure." *Annual Review of Anthropology* 42:1, 327–43.
Lewis, E. (2022). Microsoft is removing emotion recognition features from its facial recognition tech. *NBC News*. https://www.nbcnews.com/tech/tech-news/microsoft-removing-emotion-recognition-features-facial-recognition-tec-rcna35087
Likens, G. E. (Ed.). (1989). *Long-Term Studies in Ecology* (1st ed.). Springer. https://link.springer.com/book/10.1007/978-1-4615-7358-6

Lin, C. and Jackson, S.J. (2023). From Bias to Repair: Error as a Site of Collaboration and Negotiation in Applied Data Science Work, *Proceedings of the ACM Conference on Human-Computer Interaction (CSCW)*, 2023.

Lindenmayer, D. B., and Likens, G. E. (2011). Losing the culture of ecology. *The Bulletin of the Ecological Society of America*, *92*(3), 245–246.

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., and Zhang, G. (2018). Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, 1–1. https://doi.org/10.1109/TKDE.2018.2876857

Lu, M. (2022). Mobile phone repair in Ghana: Comparing three approaches. *New Media & Society*, 14614448221083652.

Madaio, M., Egede, L., Subramonyam, H., Wortman Vaughan, J., and Wallach, H. (2022). Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1–26. https://doi.org/10.1145/3512899

Mattern, S. (2017). The big data of ice, rocks, soils, and sediments. *Places Journal*.

Mimno, D. (2014). "Data carpentry," available at: http://www.mimno.org/articles/carpentry/

Monserrate, S. G. (2022). "The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage." *MIT Case Studies in Social and Ethical Responsibilities of Computing*, no. Winter 2022 (January). https://doi.org/10.21428/2c646de5.031d4553.

Monteiro, E. (2023). *Digital Oil: Machineries of Knowing*. MIT Press.

Mosco, V. (2015). *To the cloud: Big data in a turbulent world*. Routledge.

Moten, F. (2003). *In the break: The aesthetics of the black radical tradition*. U of Minnesota Press.

Muller, M., Feinberg, M., George, T., Jackson, S. J., John, B. E., Kery, M. B., and Passi, S. (2019). Human-Centered Study of Data Science Work Practices. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–8. https://doi.org/10.1145/3290607.3299018

Munoz, J. (2009). *Cruising utopia: The there and then of queer theory*. New York: New York University Press.

Murphy, M. (2017). Alterlife and decolonial chemical relations. *Cultural Anthropology*, *32*(4), 494–503.

Nemer, D., & Chirumamilla, P. (2019). Living in the broken city infrastructural inequity, uncertainty, and the materiality of the digital in Brazil. *DigitalSTS. A Field Guide for Science & Technology Studies*, 221–239.

Oladele, S. (2022, August 11). A Comprehensive Guide on How to Monitor your Models in Production. *Neptune.Ai Blog: MLOps*. https://neptune.ai/blog/how-to-monitor-your-models-in-production-guide

Palmer, P. (2010). *Domesticity and dirt: Housewives and domestic servants in the United States, 1920-1945*. Temple University Press.

Passi, S., and Jackson, S.J. (2017). Data vision: Learning to see through algorithmic abstraction. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2436–2447.

Passi, S., & Jackson, S.J. (2018). Trust in Data Science: Collaboration, Translation and Accountability in Corporate Data Science Projects. *Proceedings of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing.*

Passi, S., and Sengers, P. (2020). Making data science systems work. *Big Data & Society*, *7*(2), 205395172093960. https://doi.org/10.1177/2053951720939605

Passi, S., and Vorvoreanu, M. (2022). *Overreliance on AI: Literature Review* (MSR-TS-2022-12). Microsoft Research. https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/

Pedram, A. (2021). *Zillow Collapse Proves AI is not Perfect: Let Me Explain.* Towards Data Science. https://towardsdatascience.com/zillow-collapse-proves-ai-is-not-perfect-let-me-explain-3bdeb03b7713

Pellow, D. N., and Park, L. S.-H. (2002). *The Silicon Valley of dreams: Environmental injustice, immigrant workers, and the high-tech global economy* (Vol. 31). NYU Press.
Pipek, V. and Wulf, V. (2009). Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology. Journal of the Association of Information Systems 10(5), 447-473.
Puig de La Bellacasa, M. (2015). Making time for soil: Technoscientific futurity and the pace of care. *Social Studies of Science*, *45*(5), 691–716.
Qadri, R. (2021). What's in a Network? Infrastructures of Mutual Aid for Digital Platform Workers during COVID-19. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–20.
Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, *5*(2), 1–13. https://doi.org/10.1177/2053951718809145
Ribes, D., and Finholt, T. (2009). The Long Now of Infrastructure: Tensions in Development. *Journal of the Association of Information Systems* 10(5), 375-398.
Ribes, D., and Jackson, S. (2013). Data Bite Man: The Work of Sustaining a Long-Term Study. In L. Gitelman (Ed.), *"Raw Data" is an Oxymoron*. MIT Press.
Rosner, D. K. (2018). *Critical fabulations: Reworking the methods and margins of design*. MIT Press.
Ruhleder, K., and Star, S. L. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.
Schmitz, M. (n.d.). *Why your Models need Maintenance*. Towards Data Science. https://towardsdatascience.com/why-your-models-need-maintenance-faff545b38a2
Shapin, S. (1989). The invisible technician. *American Scientist*, *77*(6), 554–563.
Singh, R. (2020a). "The Living Dead": Orphaning in Aadhaar-enabled Distribution of Welfare Pensions in Rajasthan. *Public*, *30*(60), 92–104. https://doi.org/10.1386/public_00008_7
Singh, R. (2020b). *Seeing like an Infrastructure: Mapping Uneven State-Citizen Relations in Aadhaar-Enabled Digital India* [Dissertation, Cornell University]. https://doi.org/10.7298/sadf-ye74
Singh, R., and Jackson, S.J. (2021). Seeing Like an Infrastructure: Low-resolution Citizens and the Aadhaar Identification Project. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 315:1-315:26. https://doi.org/10.1145/3476056
Star, S. L., and Bowker, G. C. (2007). Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology*, *9*(4), 273–280. https://doi.org/10.1007/s10676-007-9141-7
Star, S. L., and Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work (CSCW)*, *8*(1), 9–30.
Tanweer, A., Aragon, C. R., Muller, M., Guha, S., Passi, S., Neff, G., and Kogan, M. (2022). Interrogating Human-centered Data Science: Taking Stock of Opportunities and Limitations. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–6. https://doi.org/10.1145/3491101.3503740
Thomer, A. K., & Rayburn, A. J. (2023). "A Patchwork of Data Systems": Quilting as an Analytic Lens and Stabilizing Practice for Knowledge Infrastructures. *Science, Technology, & Human Values*, *0*(0). https://doi.org/10.1177/01622439231175535
Ureta, S. (2015). *Assembling Policy: Transantiago, Human Devices, and the Dream of a World-Class Society*. MIT Press.
Verre, J. (2019, August 9). *Life of a model after deployment*. Towards Data Science. https://medium.com/towards-data-science/life-of-a-model-after-deployment-bae52eb83b75
Verre, J. (2020). *Monitoring Machine Learning Models*. Towards Data Science. https://medium.com/towards-data-science/monitoring-machine-learning-models-62d5833c7ecc
Vinsel, L., and Russell, A. L. (2020). *The innovation delusion: How our obsession with the new has disrupted the work that matters most*. Currency.