

## ORIGINAL ARTICLE

# Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science

Gina Neff,<sup>1,\*</sup> Anissa Tanweer,<sup>2</sup> Brittany Fiore-Gartland,<sup>3</sup> and Laura Osburn<sup>4</sup>

### Abstract

What would data science look like if its key critics were engaged to help improve it, and how might critiques of data science improve with an approach that considers the day-to-day practices of data science? This article argues for scholars to bridge the conversations that seek to critique data science and those that seek to advance data science practice to identify and create the social and organizational arrangements necessary for a more ethical data science. We summarize four critiques that are commonly made in critical data studies: data are inherently interpretive, data are inextricable from context, data are mediated through the sociomaterial arrangements that produce them, and data serve as a medium for the negotiation and communication of values. We present qualitative research with academic data scientists, “data for good” projects, and specialized cross-disciplinary engineering teams to show evidence of these critiques in the day-to-day experience of data scientists as they acknowledge and grapple with the complexities of their work. Using ethnographic vignettes from two large multiresearcher field sites, we develop a set of concepts for analyzing and advancing the practice of data science and improving critical data studies, including (1) communication is central to the data science endeavor; (2) making sense of data is a collective process; (3) data are starting, not end points, and (4) data are sets of stories. We conclude with two calls to action for researchers and practitioners in data science and critical data studies alike. First, creating opportunities for bringing social scientific and humanistic expertise into data science practice simultaneously will advance both data science and critical data studies. Second, practitioners should leverage the insights from critical data studies to build new kinds of organizational arrangements, which we argue will help advance a more ethical data science. Engaging the insights of critical data studies will improve data science. Careful attention to the practices of data science will improve scholarly critiques. Genuine collaborative conversations between these different communities will help push for more ethical, and better, ways of knowing in increasingly datum-saturated societies.

**Keywords:** critical data studies; data for good; data science; ethics; qualitative methods; theory

### Introduction

What would data science look like if its key critics were engaged to help improve it, and how might critiques of data science improve by considering the day-to-day practices of data science? Could the result be scholarship that is based on engagement with the practices emerging in data science and pragmatic solutions for creating the social and organizational arrangements

necessary to support more ethical data science? Our research on the everyday practices of data science shows the amount of reflection, guesswork, second guessing, “sensemaking,”<sup>1</sup> community building, and compromise that goes into the work of data science. Although subject to powerful rhetoric about the potential transformation of society through new kinds of data and analyses, the data scientists in our studies often recognize

<sup>1</sup>Oxford Internet Institute, University of Oxford, Oxford, United Kingdom.

<sup>2</sup>Department of Communication, University of Washington, Seattle, Washington.

<sup>3</sup>e-Science Institute, University of Washington, Seattle, Washington.

<sup>4</sup>Department of Construction Management, University of Washington, Seattle, Washington.

\*Address correspondence to: Gina Neff, Oxford Internet Institute, University of Oxford, 1 Street Giles, Oxford OX1 3JS, United Kingdom, E-mail: gina.neff@oii.ox.ac.uk

the complexity of producing, exchanging, and making sense of data, even if they do not always articulate the same kind of public critique of these practices and processes that social science and humanities scholars do.

Those critiques form the basis of a field that has been called “critical data studies,” emerging scholarship on the role of data in social life that resides at the intersection of science and technology studies, social science, policy and legal fields, and the humanities. Critical data studies claim to encompass “the types of research that interrogate all forms of potentially depoliticized data science and to track the ways in which data are generated and curated, and how they permeate and exert power on all manner of forms of life.”<sup>2</sup> These critiques point to the fact that data and algorithms are not inherently objective or fair, but rather have the potential to shape or magnify biases in judgment or processes of societal discrimination.<sup>3–9</sup> More scholars from a wide range of disciplines are turning to critical data studies, and in doing so, they are calling into question how social processes, human biases, contexts, technological infrastructures, and differences in computational tools continue to influence knowledge in an increasingly data-driven world.

For many people working in data science, *data ethics* has become a shorthand phrase for such questions—a way of signaling concerns to practitioners and critics alike. Yet the term data ethics represents a somewhat narrow placeholder for conversations that could potentially improve our ability to design research and scientific environments, and advance our understanding of society. Such a broader view of data science ethics could be based on a clear understanding of the practices of data science, while addressing the main obstacles that individuals face when trying to make ethical choices and enact them into practice.

Some of the most blustery rhetoric about so-called big data portrays the world as knowable through algorithmic decision-making logic, data as independent of the infrastructures that helped produce them, and “truth” as outside social processes of negotiation, human understanding, or theory.<sup>10–12</sup> Often data science critiques are aimed at such rhetoric. However, most of our observations of data scientists showed the messy work of people consciously wrestling with exactly these issues and tradeoffs. For data scientists, the contingency of their data and the social implications of their analyses were often very much a part of their daily conversations and technical choices. Still, the work of ethically motivated individual data scientists too often does not ad-

dress the broader ethical implications of their work on organizational, institutional, and societal scales.

Our aim in this article is not to distinguish between reflexive and unreflexive individual data scientists and their everyday practices. Rather, we hope to address how to support the integration of the perspectives of critical data studies into the organizational and social processes fundamental to doing data science. As ethnographers observing the social and organizational work of data science, we saw these negotiations in practice, and we attended to how teams are already addressing many issues raised by critics. How to build ethical data science is far from a solved problem, and data science teams continue to grapple with it. By understanding and articulating the social and organizational work inherent in making data and in doing data science, we hope to outline the opportunities for integrating the language and experience of data science into its critique, and in the process help to advance more ethical practices in data science that expand on the valid critiques of those outside it.

To do this, we present a set of methodological and theoretical considerations for analyzing the practices of data science. Our research focuses on the complexities of the grounded practices of making and exchanging data and of the processes of communicating about data across professional and organizational differences. We look at how data scientists work within teams, at their understanding of their jobs, and at the constraints of the settings where they work. We were immersed in and working with data-making and data-analyzing communities in academic data science, “data for good” projects, and specialized cross-disciplinary engineering teams. The data scientists we observed frequently understood and advocated for considering competing values and ethical choices in their work, and were more reflexive about their practices than critical data studies often present them as being. Our position straddling the boundaries between the conversations critiquing data science and advancing it means that we see the lessons that could be shared for both purposes.

With that in mind, we outline four key components of data study critiques and how they relate to our observations of data science in practice. These include (1) data as inherently based on interpretation, (2) data as reliant on social context for meaning, (3) data as the product of the social–technical arrangements that produce it, and (4) data as the media through which conversation, negotiation, and action can occur. We show how each of these propositions from critical data studies

played out in the day-to-day practices of data production, data analysis, and data-driven decision making that we observed. We do this, in part, to bring closer together the work of data scientists and scholarly conversations on critical data studies.

We cannot possibly hope with one article to resolve differences in ways of knowing among different kinds of data scientists and between data scientists and their critics, at times stark ontological differences in understanding what data are and what data can do. Rather, we want to show what others might learn from the practices of data science, how these observations might improve critique, and how to integrate critique into the day-to-day practice of data science. Our research with different data science teams struggling to make sense of data, communicate with data, and produce data in teams and groups shows the complicated ways that data deviate in practice from ideals of pure scientific objectivity or bias-free representations of society. The practices of the teams we studied suggest pragmatic ways that critics and data scientists through collaboration can enrich each other's research.

### Methods and Settings

Two authors (Tanweer and Fiore-Gartland) have spent more than 2 years doing ethnographic research on the practice and culture of data science in U.S. academic settings. Our research engagement includes Tanweer's and Fiore-Gartland's employment within one of the largest academic data science initiatives in the United States, a capacity in which we have participated in strategy and planning meetings, interviewed and surveyed participants in collaborative and educational data science programs, and conducted close ethnographic observations of the daily work of the initiative. Our participant-observer role is based on a growing recognition that social scientists and humanities scholars positioned within research teams can "offer both specific recommendations and a knowledgeable perspective from which to weigh social alongside technical and scientific concerns" (p.1),<sup>13</sup> taking an active role in improving scientific practice.<sup>14</sup>

Our team of ethnographers, funded by and embedded within the data science initiative, has had the opportunity to observe, participate in, and shape how data science programs, organizational processes, and practices evolve. A major focus of this ethnographic research is on the work involved in communicating across sectors and disciplines,<sup>15</sup> developing emergent models of education and organizing,<sup>16</sup> and in seeing and knowing the na-

ture of increasingly large, distributed, and heterogeneous datasets.<sup>17</sup> At the time of writing, this research has produced 100 interviews and roughly 2450 pages of field notes and in-process analytic memos from which we drew the ethnographic vignettes used in this article.

The other two authors on this article, Neff and Osburn, are part of a larger research team focusing on how results of energy modeling and energy engineering are translated and synthesized within architectural design and building teams.<sup>18,19</sup> We observed three energy modelers working on 18 different building projects and conducted over 313 hours of field observations of energy modelers at work, resulting in 297,753 words of field notes. Our ethnographic observation team consisted of three researchers, one each observing work at the energy modeling office and the project's architecture firm and one participating as a design team consultant. In addition to researchers being both participants in and observers of the team making the engineering analysis, we also interviewed, in depth, seven of the key people on the particular project we discuss in this article. The vignettes from the architectural design project that we present in this article comprise just a small portion of our qualitative data analysis of how communication practices shape data practices within teams.

### Data as Interpretation: Harnessing Data for Social Good

A key insight of critical data studies is that interpretation is cooked into the very structures of data and that the work of claiming something as data or a dataset becomes a "rhetorical" move. *Data*, as a word, although ends up sounding more authoritative than perhaps those who produce it ever intended:

At first glance data are apparently before the fact, they are the starting point for what we know, who we are, and how we communicate. This shared sense of starting with data often leads to an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself (p.2).<sup>20</sup>

Critical data study scholars question whether data can ever be truly objective. The *data as interpretation* critique sees human judgment not as a contaminant that can be removed from data, but a reflection of the inherent choices that go into the construction of data sets, which necessarily rely on patterns of inclusion and exclusion. Raw data are, thus, both "an oxymoron and a bad idea" (p.184).<sup>21</sup> the social is always implicated within data even if people producing data and analysis do not describe it as such. Implicit in

scientific data sets are processes of counting that can be understood as “epistemic achievements that involve categorical judgments” (p.246).<sup>22</sup> These processes are not restricted to so-called social data, or data generated from the traces of behaviors and activities of people. Every disciplinary institution and body of knowledge has “its own norms and standards for the imagination of data,” and “different data sets harbor the interpretive structures of their own imagining” (p.3).<sup>20</sup> In other words, the context in which people work leads them to make different choices and definitions for the data they work with. When people share data across contexts, they do not necessarily include those “imagined” or taken-for-granted norms and standards of data from their context.

Data are further interpreted as they are “cleaned,”<sup>4</sup> a process that involves subjective judgment about what data are good, what data count, and how data should be transformed. Even the acts of framing what questions to ask and choosing techniques from among an ever expanding suite of analytic methods involve judgment calls informed by a data analyst’s identity and situated perspective.<sup>4</sup> This is true for how different stakeholder groups interpret data and what expectations they fostered for how data are produced and used.<sup>23</sup>

In the various “data for good” spaces that Tanweer and Fiore-Gartland have observed, most of the people involved with the work of data science acknowledge to some extent the contingent, constructed, and value-laden nature of data. This often arises very starkly when data collected for one purpose is applied to another purpose, an act of translation that has become a signature move in data science.<sup>24</sup> A case in point is when administrative and transactional records from the public sector are made available for “social good” purposes.

For example, one project we observed was built around a transit agency’s electronic payment information, data that were generated when customers paid for their transit usage by “tapping” an agency-issued prepaid card while riding local buses and trains. In the past, these data had been used for transactional purposes such as monitoring riders’ account balances and for operational purposes such as performance evaluation and management. However, the team of academic researchers we were observing set out to demonstrate the broader utility of analyzing these data, hoping to prove its usefulness for making predictions and strategic planning. Their labor demonstrated that the value of these data was not self-evident and re-

quired a deeper, forensic look at what might represent. One researcher remarked that the fundamental challenge they faced was that the system generating the data “was designed as a payment system, not a data system.” This lay at the heart of what became the focus of the research team, to disentangle the assumptions and values that were baked into the data at their creation and to reveal their biases. The team decided the questions that most interested them revolved around equity of services across different communities. However, in trying to ask these questions of a dataset that was collected to record payments, it became difficult to disentangle what communities and patterns of use the data represented. In this dataset “ridership” was interpreted as those customers who paid using the electronic payment system and did not include those customers who paid with cash. The team realized it was possible that low-income transit users were more likely to make cash payments. The extent to which users of the electronic payment system were representative of the city’s demographic composition was unknown and represented one of the possible biases in the data. One researcher described their purpose as “making the data less terrible” by giving transit agencies an idea of where the data were incomplete and “who’s not in the data.” They were concerned that their analyses and any future analyses relying on these limited data might amplify existing biases. In attempting to use this public transportation transactional data to answer research questions, they were confronted with having to understand the interpretive frameworks within which the data were created in the first place.

Recognizing how data are constructed and the ways in which they harbor interpretations, as this team did, is essential to surfacing the limitations and possibilities of future analyses and uses of data. The project leader believed wholeheartedly in the abundant value of these data, but also recognized that the kinds of interpretations they could make “were constrained by the data we have.” Even as “big data” are promoted as rich raw resources from which all sorts of value may be derived, data science practitioners are frequently confronted with a less-than-complete understanding of the assumptions and values shaping the data they are working with. Figuring out how to bring their conversations out into the open with more and different stakeholders and how to reproduce practices of reflection in other data science settings are how the *data-as-interpretation* critique can lead to actionable recommendations for creating a more conscientious, circumspect, and ethical data science practice.

### Data as Context: The Challenges of Reproducibility in Academic Data Science

Critical data study scholars warn that when context is not appropriately considered, “data lose meaning and value” (p.670).<sup>4</sup> For example, interpretations of data drawn from social media sometimes neglect to account for the differences between the population of users of that platform and the general population, missing the interaction of social media data and the population.<sup>5,25,26</sup>

While lack of contextual understanding is often at the center of big data critiques, context is also a central concept for the work of human computer interaction, ubiquitous computing, and data science.<sup>27,28</sup> Seaver’s study of “context-aware” music recommender systems revealed that for developers of these systems, operationalizing context is key for the goals of personalized recommender systems that aim to adapt their recommendations to who you are, where you are, and when it is.<sup>28</sup> As a concept, then, context often matters greatly to both critics and computationalists like Seaver’s developers and data scientists. The issue of whether or not context is important is moot. Instead, given that “the nice thing about context is that everyone has it,”<sup>28</sup> Seaver suggests it might be more instructive to think in terms of “context cultures” to make sense of which context matters when and to whom.

Context has become an important and prevalent term across data science practitioners and critics, providing a common point of reference for the everyday practice of data science, even without a shared understanding around what it is and the cultures that produce it. Seaver draws on Dourish’s<sup>27</sup> distinction between representational and interactional context to illustrate two differing views of context and how it is that “context can be simultaneously missing from data science and central to it” (p.1103).<sup>28</sup> Computationalists often adopt a representational view of context when they imagine that context can be represented as data, or as a “stable container” within which activities unfold (p.1105).<sup>28</sup> This view imagines context as an environment with definable, describable, and encodable boundaries and characteristics that exist outside the data.

We observed this representational perspective of context operating across the discourse and practice of data sharing and reproducibility in data science. Making data science reproducible entails not just publishing how one achieved particular results, but providing the broader context or environment in which those results were achieved, opening datasets, codebooks, code, documentation, and coding platforms, to others so that they can replicate results, reproduce the entire study, or extend the work in

new directions. Data scientists often implicitly or explicitly assume encoding more “context” around data can solve many challenges of sharing data, data mining, and reproducibility in data science. We observed how the call for reproducible data science in academia often imagines computational research as a suite of technologies and data with a representational context that can be neatly defined, wrapped up, packaged, and transported to be used by someone else.<sup>29</sup>

However, this perspective overlooks thornier questions of what and how context is accounted for, how context is constructed in the moment as a local accomplishment of the process of data science, and how context is deeply entangled in problems of communication, issues often taken up by critical data scholars. From this perspective, data are always already context-rich because of how people imagine data and construct, produce, or define the dataset. These relational properties of data, as Borgman observed, mean “what are data to one researcher are context to another” (p.518).<sup>30</sup> These ideas support what has been called an interactional view of context, seeing context as “relational properties occasioned through activity”(p.5),<sup>28</sup> in which data are not separate from practice. To critical data scholars, context does not preexist, but is instead a local accomplishment that can shift dynamically and can never be wholly captured as data.<sup>27</sup> Data scientists in our field site made data reuse and sharing possible through their conversations and interactions that brought relevant context to the data. It was not enough to rely on context encoded as metadata or a solely representational view of context.

People used the term *context* as if it were one thing, yet simultaneously meant very different things by it, and in practice we found the representational and interactional views entangled and inextricable from the other. These different notions of context became very clear to us while observing the project of an oceanographer named Rachel as she worked with a dataset compiled and synthesized from dozens of distinct ocean expeditions. As she tried to analyze these data, she encountered a very abrupt and unexpected jump in her results. This was an example of what many in the field call blockers, one that often triggers a more granular look at the datasets to learn something about one’s data.<sup>17</sup> In looking more closely, she realized that one cruise’s measurements were drastically different from the others. Rachel, then, investigated what was at the root of this difference, so she could adjust her analysis and interpretation accordingly, but the answer did not exist within the dataset. Rachel tracked down another

researcher who was on the cruise in question to ask why the measurements were so different. The other researcher explained that the instrument sensitivity had been adjusted on that cruise, information that had not seemed relevant to record and share before knowing that the data across multiple cruises would be synthesized and analyzed in this way.

Once Rachel had uncovered this important contextual information about the data's provenance, she could calibrate the measurements that were collected using one level of sensitivity with the measurements that were collected using a different level of sensitivity. However, she also recognized that these kinds of instrumentation adjustments were something that recur given how data were collected. Rachel decided to try prevent the same thing from happening to other researchers in the future, which required working with personnel in charge of instrumentation to open channels of communication and make sure that changes to the instrument settings on cruises would be automatically recorded and time-stamped as the data are collected.

Rachel's story illustrates the entanglement of representational and interactional views of context. Initially Rachel is in search of relevant context that could help her understand the inconsistencies she encountered in her compiled dataset. In practice, this was not simply obtaining the relevant metadata about instrumentation adjustments as a representational view would suggest. In fact, the process of obtaining this information required she establishes social connections with other researchers and together figure out what constituted "relevant" metadata. This context is not self-evidently relevant, and only emerges as relevant, as context, in relation to a situation or activity. This is interactional context. In practice, we see how interactional and representational context are intertwined, and in this case, one is inextricable from the other. To help future researchers make sense of similar data, Rachel had to translate across "context cultures," establishing new channels of communication as well as developing a process for encoding relevant metadata into future data collection.

Just as most would agree that data should not be taken out of context, critical data scholars need not take data science out of context by considering analytical products or artifacts separately from the work that constructs them. What had appeared to be a problem of representational context for Rachel in her effort to repair relevant metadata, in fact, required human interaction through which relevant context could be constructed and emergent commu-

nication pathways could be established. This case and many others like it suggest that an expanded and integrated notion of context would move us toward better supporting the pathways for interaction and communication around data. This is work that we already see evidence of in data science practice, but is not often recognized as such. Echoing how Dourish<sup>27</sup> positioned the overlapping cultures of context, we argue that such attention could help resolve what may otherwise appear as incompatible views between practitioners and critics.

### Data as Mediated: Design Teams and the Making of Engineering Analysis

Critical data studies show data as being produced by and inextricable from the "software, hardware, instrumentation, protocols, and documentation" that helped to produce it.<sup>31</sup> This critique holds that data are a product of the *sociomateriality* of their production—the methods, instruments, and other technologies, but also the discussions, negotiations, and assumptions that happen while making data and analyses. Data entail both work, or "information labor,"<sup>32</sup> and large social and technical systems for data production, exchange, circulation, and maintenance.<sup>33–38</sup> "Data ecologies,"<sup>39</sup> "data assemblages,"<sup>40</sup> and "knowledge infrastructures"<sup>34</sup> shape what questions can be asked, what data are sought, and what analyses can be thought of. These systems comprising people and technologies can be said to *mediate* data or are, in other words, the medium for data.

Consider the teams of engineers and architects that we (Neff and Osburn) studied. They used predictive modeling of energy consumption to create frameworks for making decisions during the architectural design of buildings. Energy modelers model, engage in guesswork, and tell stories about and with their data, while interacting in interdisciplinary and interorganizational design teams. Energy modelers in the United States typically use the same sophisticated predictive modeling software that was produced by the Department of Energy. Each team we studied was keenly aware of the complexity of the industry's standard computational platform, and yet the computational analysis required specific recommendations and professional judgments for the initial assumptions needed to build a model. Energy modelers in interviews said estimating or guessing was a key part of the expertise needed for their jobs. In essence, the computational modeling tools were fairly standard, and the role of expert energy modelers was to fit the data to the particular situation at hand and then to figure out how to communicate the data and results to people outside

the process. Energy modelers understood that their key energy performance metric, EUI (energy use index), was partly a product of software standardized across the industry. They also saw their professional role as helping the design team construct the data and analysis by guiding the assumptions and choices used in an energy model and negotiating with the design team on how to present the model and the design consequences it implied to clients.

The energy engineers we observed worked on the design of a new hospital building. From the beginning, the architectural design team wanted to use modeling to help the hospital's board compare different designs based on their energy use, cost, and potential energy savings. However, the hospital's owners had selected a seemingly ambitious EUI goal for the project, which suggested to the architect that they valued sustainability and might be persuaded to pursue a more aggressive energy goal than the initial target. Other information from the hospital's owners, however, contradicted that this level of energy efficiency was the priority over lower construction costs.

The ambiguity over what the owners wanted led the design team to debate what assumptions should go into the energy model. At the crux of these debates were stories the design team told each other around how the owners understood energy modeling data. Were they experts, clear about energy efficiency goals and the meaning of the index numbers that they had requested as targets and would be produced by the model? Or did they arbitrarily choose an EUI with little awareness of what was required to achieve it? Would they be interested in high performing, but expensive systems? During these discussions, the design team recognized the energy predictive modeling data as contingent on a whole host of variables that they had control over, but also dependent on the unknown meaning behind the owner's energy goal. This meant that project team members spent much of their time in guesswork and negotiation: they searched for "hints" about the owner's priorities and negotiated over how to interpret these "hints" into assumptions to make the model.

The energy modelers saw the data they worked with as the result of both the methods of production—algorithms, software, and hardware—and social processes consisting of negotiated "guesswork" and stories. Energy modelers said in interviews that the model assumptions were developed based on extensive expertise and team negotiation about how to make and use the data to present a specific story about design options that would support decision-

making. Still, at many points, while producing the energy model, they would defer to how a particular set of options would be "modeled," moving between a best guesstimate of the EUI those options might produce and the EUI generated by the software. This is not bad science or engineering. On the contrary, it is actual science and engineering how it is practiced every day.

The social processes of storytelling, guesswork, and negotiation helped the design team with "sensemaking"<sup>1,41</sup> around who the data are for, what the data need to do, how the data need to be made, and how data can persuade. These social processes in turn shape the narratives that data can tell and influence how data are interpreted. Understanding the social and technological processes of mediation can help data scientists make sense of what to do with data and give sense to the data for others. Having a critical data study scholar at the table can help data scientists make sense of the potential array of meanings behind ambiguous goals, ideas, or challenges that their data are intended to achieve or resolve, and the implications of these interpretations. Likewise, watching data scientists in action as data are mediated provides insight into how data scientists are cognizant of and negotiate with other team members and stakeholders around the social choices and tradeoffs that are inherent to data production and modeling.

### **Data as Media and the Process of Team Negotiations for OpenStreetMap**

Data can serve as media for moments of connection, conversation, meaning, and other social activity. Critical data studies show how data take on other social functions in addition to the quantification of an empirically knowable reality. Numbers, for example, can "mediate between and across various realms of meaning and knowledge," helping express personal and private experiences so that they can be compared.<sup>42</sup> The data become a medium for thinking about collections of items, observations, actions, or people—from datum to data. The media function of data means that people can "*constitute and enact* their relations with one another *through* the use and exchange of data" (p.75).<sup>43</sup> Data can become the impetus for a conversation or an opportunity to make a connection<sup>23</sup> and can offer "voice" and other forms of expression.<sup>44</sup> Data "may be contested at the boundaries of institutions and communities" (p.1480),<sup>23</sup> leading to the development of approaches for translating people's varied values and expectations for using data. Data are also seen as a medium through which other social values or intentions

can be communicated, constituting a “data citizen,”<sup>45</sup> “calculated publics,”<sup>6</sup> and “computational politics.”<sup>26</sup>

In the course of our fieldwork, we (Tanweer and Fiore-Gartland) have seen practitioners acknowledging this critique, while intentionally and reflexively instantiating their values into code. This was the case in one “data for good” project we observed that had the ultimate goal of building a routing application for people who use assisted mobility devices such as wheelchairs. For the base layer of their app, the project team hoped to use OpenStreetMap (OSM), an open-source, user-generated mapping platform. However, sidewalks and curb cuts, critical features for people with limited mobility, were documented in OSM only as attributes of streets, rather than as distinct features in their own right. This posed a number of problems for the team because it meant that a routing algorithm could not map a course along a sidewalk, and it made the visualization of sidewalks and curb cuts on a map quite difficult. The OSM data did not work for the team’s purposes, and so they decided to try to change the OSM standards.

Much of the team’s work toward developing that new standard involved deliberations about what data should be generated, how it should be represented, and how the pragmatic and ethical rationale for those choices should be communicated. Building a shared data platform became the medium through which such conversations happened. These conversations took place both internally among the team members and with other stakeholder groups, making their work look more like advocacy and political mobilization, although for data standards. The team interviewed users of assisted mobility devices to find out what kinds of information were useful to them in navigating the city. They studied previous attempts to change OSM standards by poring over hundreds of pages of discussion threads and listserv archives to ascertain what the concerns of the community were and what approaches could lead to acceptance. They sought out leaders in the local OSM community to get their advice on how to proceed and they presented their ideas to national and international audiences through the community’s established channels of communication.

This process of negotiation about and through the data produced many things that influenced the team’s work. Data became both the occasion for, and the culmination of, a complex series of negotiations, arguments, and valuations. For example, they learned that some members of the OSM community were concerned

about separating sidewalks from named streets because it could make turn-by-turn routing directions more difficult. They found that some wheelchair users did not worry much about curb cuts, but others did. Also, they realized that, while focusing on the needs of people with limited mobility, they had not been considering the distinct needs of people with impaired vision. At each decision point brought on by these realizations, they reflexively discussed how their options entailed making choices about whose experiences and perspectives would be validated and represented in the data standards they ultimately proposed, with full recognition that the data would become a mechanism of conveyance for subjective values and priorities.

To be sure, compared to the team working on introducing new OSM standards, not all the data-intensive projects we have observed were quite as conscientious and systematic about using data as an occasion for sense-making of multiple stakeholder perspectives or as aware of how data become a vehicle for the expression of values and priorities. The level of awareness and reflexivity of the OSM team may not be the norm. However, we would argue that critical data scholars and data science practitioners alike have much to learn from the situations in which data are productively and ethically leveraged as a medium for complex stakeholder negotiations. We can use examples like these to identify and foreground the set of skills that are critical to a principled and inclusive data science practice, including the ability to engage with diverse stakeholders, to recognize multiple, potentially conflicting needs, and to incorporate multiple perspectives into decisions made about and with data.

### **Tools for the Future of Data Science (and Its Critiques)**

In each of our cases, we observed nuanced, contextualized, and reflexive practices of individual data scientists at work. To be fair, our cases, environmental energy engineering and academic data science, may be somewhat insulated from forces that could make ethical choices about data more difficult in other settings. The teams we studied have a high degree of autonomy, which may not be the case for other types of data science teams. Much of the data science work in academia, and the data for good projects in particular, are far removed from profit motives at the heart of most commercial data science. However, in these four settings, we saw people acknowledge their own interpretive contributions, consider the context in which data scientists’ work is enmeshed, recognize the way their work



is mediated by sociomaterial processes, and use data to surface and negotiate social values. In the day-to-day practice of making and doing data science in these settings, we saw little evidence of the unreflexive data scientist, unlike some portrayals in critical data studies. Our work suggests that critiques become apparent to many individual data scientists and teams as complexities arise in the course of their work. This was also the case in a study conducted by Hautea et al.,<sup>46</sup> when they invited youth participating in the online community Scratch to analyze public data generated about their own learning and social interactions within the community. The youth recognized and articulated a number of ethical concerns that authors call “critical data literacies,” and some expressed a heightened concern for privacy given the ability to conduct surveillance programmatically, questioned the assumptions underlying their analyses, and wondered if their decisions about which interactions to count and analyze would, in turn, shape the kind of interactions the community valued.<sup>46</sup>

The lesson to be drawn is neither that the work of critical data studies is done nor that critics are wrong in their assessments of unethical or unreflexive work. Rather, we think that by learning from examples of best practices for a socially informed, or “human-centered,”<sup>47</sup> data science, we can improve how data science is done. Ethnographic research of the day-to-day practices of making data and doing data science, like we and others are doing, can help us think with, not for or against, the engaged practitioners of data science as they reflect on their work, to inform how to design new kinds of ethical institutions.

One reason for this work is that organizations may fail to make ethical decisions even if ethical people are involved. danah boyd warns us against overlooking the “messiness” of ethical thinking in complex organizations and technical systems: “How do we enable ethics in the complex big data systems that are situated within organizations, influenced by diverse intentions and motivations, shaped by politics and organizational logics, complicated by issues of power and control?”<sup>48</sup> That remains, in large part, the question to ask. There is an urgent need for organizational cultures, practices, and structures to support the deliberation of ethics in data science; so that, instead of individuals grappling with ethical dilemmas on an *ad hoc* basis, we create robustly ethical institutions that consider how the social is implicated in data at every step. Such organizational cultures would empower people to act on their concerns and support organizational processes to help reflection

to be incorporated into decisions concerning data and analyses.

Below we present a set of four methodological and theoretical provocations to support a more informed, robust, and ethical data science practice across varied settings. It is our hope that such work brings together conversations between data scientists and critical data study scholars to incorporate the best of social science and humanistic scholarship into the day-to-day practices of making data.

#### (1) Communication is central to the data science endeavor

The emerging field of critical data study tells us that context matters to the understanding of data and data themselves serve as media for negotiating meaning. The field research we presented in this study show the importance of communicating context and the impossibility of separating data from their context. While communication is often considered important for the conveying of results or distribution of end products, we see people thinking about the process and value of scientific communication as beginning with data making, data processing, and data analysis, not just as a final step in the form of publishing. Methods for studying communication can be applied to the practices of generating data, talking about data, and organizing around data.<sup>49</sup> Our research suggests that data science practitioners should think about the entire process of data gathering and production as one that has communication at its core and communication practices as key data science practices. This means that data scientists should think of all data as partially “social,” not simply because of social function or origin of data, but because the social processes of making data are acutely important to the outcome. Data science practitioners need to consider communication before and during data production and how their communication practices shape how data are translated across multiple audiences and contexts because this is at the root of their science.

While we, four communication scholars, readily recognize that communication is already playing a central role in data science, we have also seen this communicative work consistently disregarded. It has become commonplace for data science practitioners to repeat the truism that 80% of data science work is cleaning the data and only 20% of the work is analysis. There is no room left in this hypothetical pie chart for the many essential layers of communicative labor that we documented. When this labor is acknowledged, it is often

represented as a distraction from the tangible work of “science.” For example, even as the team working on the proposed OSM standard understood that their painstaking deliberation and outreach were crucial to the success of their project, a couple of the team members were sullen about having to spend so much time “just talking,” discouraged at their lack of “progress,” and anxious to get to the “work” of writing code. Their experience shows that the practices of data science are fundamentally also about social networking and organizing, about having conversations among colleagues, and building social relationships to produce meaningful datasets.

Even though communication plays such a central role in every step of the data science process, it is marginalized in terms of what counts as work and what gets valued as a contribution. We need to find ways of rectifying this to recognize communication as an integral part of the data science endeavor.

#### (2) Making sense of data is a collective process

Critical data scholarship has also taught us that layers of interpretation are built into every step of making and using data. Data science is no different. Rather than seeing interpretation and analysis as the product of individuals, consider how sensemaking takes multiple people. For the engineering teams we studied, knowing what should count in the model was a process of negotiation that was inseparable from the final model outcome. Knowing why the model produced a “better” outcome than expected also depended on collective expertise of both the architects and engineers on the project. However, the knowledge that model making and sensemaking of energy data were a fundamentally social and collective process was not clear outside the meetings between the energy engineers and architects, and, in particular, with the hospital’s owner. The sensemaking process of data science requires work because the connections and relationships among the tools, data, and analysis are not self-evident.

Acknowledging that sensemaking is always a collective process means it is imperative to support practices that draw relevant voices into conversations around the production and use of data. For example, advancing reproducible data science necessitates not just version control and thorough documentation to enable sharing data and code but also accessible channels of interpersonal communication and norms of cross-disciplinary cooperation. Also, any study or project that involves the use of social data should involve careful stakeholder analysis to consider the multiple perspectives that are implicated or impacted by the use of such data.

#### (3) Data as a starting point, not end

Our research shows that data have multiple potential pathways in its production. The work of making and analyzing data is a journey, not a destination, the product of layers of contributions from multiple people; so data often lead to new questions. Making sure to bring people into this process is the starting point of “open science,” which was the goal of one of our teams. However, it is also the starting point for a whole host of political engagements with data. For example, grassroots communities and patient advocates often want to be part of making data, rather than having data done to them and analysis performed about them.<sup>23,50,51</sup>

Data can also be a “site for conversations”<sup>23</sup> before, during, and after the production and exchange of data. Using data as the opportunity to make transparent the assumptions and deliberations that go into choices and to ask more questions, get more input, and build even richer “context.” These activities should be seen as central, not peripheral, to data.

#### (4) Data as sets of stories

That data are made from and exchanged through sets of stories is clear in each of our field sites. Stories occur before data production, during production, and are used in exchange to give data meaning across communities with different expertise, cultures, and practices. Communication shapes data and molds data’s rhetorical possibilities. This would seem to have different, but equally important, applications across the different contexts. For engineering, the stories that data can tell begin with the stories that shape the production of data or the stories that help make sense of the potential desired outcome and need for data. Data practitioners, organizations, and other stakeholders should listen for stories, hints, and assumptions about the intended audience for data. These can be transformed into sets of stories that merge into a larger narrative that shapes how the data will be produced and what people expect the data to do. Through listening for such hints, the use of data in the larger narrative can be framed in a way that aligns with the intended audience’s own goals and values and can help lead or persuade other stakeholders toward specific choices in decision-making. Data are, as seen above, rhetorical. If data are to be used for social good, for example, then the rhetoric of social good must be implicated before and during the production of data to lead toward ethical and socially responsible data-driven decision-making.

### Conclusion: Pragmatic Steps Forward

We have identified two broad categories of actions that critical data study scholars and data scientists can take for advancing a data science that is sensitive to deep social implications of the nature of data. At the heart of these actions is the need for scholars to use their social science and humanities expertise and apply methodological and theoretical tools that center on communication, sensemaking, mediation, and organizing. These actions will help data scientists better understand and articulate the potential implications of their practices that can lead to integrated ethical thinking within data science across more settings.

- (1) Create new opportunities for data science practitioners and data science scholars to engage in sensemaking together

The first action is to continue to bring together those who practice data science with the social science and humanistic scholars who critique the ethical, political, and cultural implications of data-driven decision making. We urge those in the latter community to get involved in observing the day-to-day practices of the work of data science and to work with data scientists who struggle with ethical choices, yes, but also with the work of collective sensemaking, communicating, and organizing. Making data meaningful across disciplinary and professional boundaries is difficult, time-consuming, and requires translation across multiple knowledge domains. Critical data study scholars have a key role in helping data science teams work out the best approaches to such translation. The best stories and frames around data can resonate with others and make data valuable, meaningful, and actionable.

Scholars who take practice-based approaches can observe disciplinary and professional discourses, practices, cultures, social norms, and social interactions to suggest interventions in real time. They can also help data science teams understand the important negotiation work that goes into the production and exchange of data and allows data to become meaningful across community boundaries and bridge conflicting obligations. Through leveraging their training and knowledge in the humanities and social sciences, scholars can help data scientists understand the layers of social, organizational, political, ethical, and emotional complexities embedded in their work, helping them to better articulate the implications of their work when making decisions around data production and exchange.

For example, in our research with a large data science initiative, the ethnography team initiated a series of conversations at the university under the heading of Data Science Studies that brings together researchers from across the university and the wider community. Participants include data scientists, social scientists, science and technology study scholars, librarians, as well as data science practitioners, and advocates from the commercial and public sectors. Many participants have noted that these meetings seemingly fill a void for such conversations, connecting those who are developing data science critiques and those who practice data science every day. The meetings center around the social and organizational dimensions of data science, including topics such as privacy, transparency, and seamless design, democratization of data science, and critical data literacies. Part of the challenge of these meetings is to find the key fusion points between critical data scholars and data science practitioners, such that ideas may be exchanged even when their epistemologies and languages might be quite different.

In our experience, data scientists have appreciated the opportunity to talk about the daily complexities of their work, including the ethical implications of certain data work decisions. Having regular conversations with scholars who are also team members can help data scientists with their own sensemaking around specific organizational, interpersonal, and communication challenges during projects, challenges that can have deeply ethical implications. In our work with energy engineers, the ethnographer at the table became a sounding board to make sense of interpersonal exchanges and conflicts, to retell and negotiate stories about the team and the project, and to think through strategies for improving communication and meaning-making around data. Sensemaking conversations such as these can help data scientists shed light on ways to overcome project challenges, consider the ethical contours of their data practices, and develop decision-making strategies that help them achieve their goals. Bringing social scholars to the work of data science recognizes these scholars not only as observers or critics but also as experts in many of the key social processes of data-making and data analysis.

Working with teams will also improve critique. We urge other scholars to join us in the messy work at the intersection of these domains. While the professional goals of the data scientist and the critical data study scholar differ, the results of increasing interactions among them will provide more robust and

nuanced critiques of data science, while improving the data science field by articulating its layered complexities and ethical implications.

- (2) Support new kinds of organizational arrangements to foster a culture of ethical data science practice

As data science evolves and matures, organizations and institutions must figure out how to incorporate it into their practice, culture, and politics. As evidenced by the initiatives we are involved in, there are currently concerted efforts underway to do this. Considering data science a thing apart from organizations, either because it is viewed as esoteric knowledge requiring highly specialized skills that only those ordained with the title “data scientist” can do or because it is viewed as purely technical expertise to be cast in a supporting role when other approaches fall short, would pose a greater challenge to robustly ethical data science practice.

As our research shows, diverse expertise needs to be brought to specific and particular problems that often have their own deeply embedded knowledge practices. This means data science cannot solve problems in a vacuum, but must be connected in real and meaningful ways to the settings where those problems arise and people who have a deep understanding of those problems. Only then can the conversations, investigations, and negotiations that we have argued above are so crucial to an ethical data science practice be surfaced, prioritized, and acted upon. The design team we discussed included architects, engineers, and energy analysts working together across very different considerations (financial, design, constructability, and hospital safety) so that their data and analyses reflected the real-world situation. In the academic setting, integrating data analytic capacities into teams changed the nature of the kinds of questions that could be asked and improved the capacities to address interactional context. When data science is integrated into existing scientific laboratories, this is true for the collective work of the laboratory, not just for those individuals who are directly working on computational tools and solutions.<sup>52</sup> This suggests that data scientists need to be mainstreamed into practice instead of being siloed as an area of specialization or shunted off into a technical infrastructure role.\* When integrated, data scientists can better understand varied ethical considerations for diverse cultures or communities and get a broader view of the social nature of data.

\*We are grateful to the reviewers and editors of this journal for helping us to address this point.

Similarly, organizations face challenges in developing pedagogical opportunities to meet both the growing demand for data science skills and the fast-paced nature of change in data science methods and technologies, all while maintaining the depth of critical reflection and sustained interactions that we have argued for. The emergent and evolving nature of technologies that support data science requires constant updating of skills, the learning of which takes significant amounts of time and often is not supported through conventional offerings. As such, we see all kinds of new opportunities forming in the interstitial spaces of institutions to facilitate the exchange of and learning about methods, data, and tools. These include offerings such as hackathons, research incubator programs, and peer-based instruction such as Software Carpentry. Institutions must figure out how to support such emergent forms of organizing in a way that promotes stability, continuity, and sagacity of the social networks that are required to mature an ethical data science practice and culture.

Engaging the insights of critical data studies will improve data science. Careful attention to the practices of data science will improve scholarly critiques. Genuine collaborative conversations between these different communities will help push for more ethical, and better, ways of knowing in increasingly data-saturated societies.

### Acknowledgments

No author has a financial conflict of interest. This material is based upon work supported by the National Science Foundation under Grant No. 1300271 “Reduce Energy Consumption through Integrated Design: How Do Engineers Translate and Teams Synthesize Energy Modeling in Successful High Performance Building Design” and the Moore and Sloan foundations.

### Author Disclosure Statement

No competing financial interests exist.

### References

1. Weick KE. Sensemaking in organizations. Thousand Oaks: Sage Publications, 1995.
2. Iliadis A, Russo F. Critical data studies: An introduction. *Big Data Soc.* 2016;1–7.
3. Barocas S, Selbst A. Big Data’s disparate impact. *Calif Law Rev.* 2016;104:671–732.
4. boyd d, Crawford K. Critical questions for big data. *Inf Commun Soc.* 2012;15:662–679.
5. Crawford K. 2013. The hidden biases in big data. *HBR Blog Netw.* Available online at <https://hbr.org/2013/04/the-hidden-biases-in-big-data> (last accessed March 29, 2017).
6. Gillespie T. The relevance of algorithms. In: Gillespie T, Boczkowski PJ, Foot KA (Eds.): *Media Technologies*, Cambridge: MIT Press, 2013, pp. 167–193.

7. Sweeney L. Discrimination in online ad delivery. *ACM Queue*. 2013;11:1–19.
8. The Leadership Conference on Civil and Human Rights. 2014. Civil rights principles for the era of big data. Available online at [www.civilrights.org/press/2014/civil-rights-principles-big-data.html?referrer=https://www.google.com](http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html?referrer=https://www.google.com) (last accessed March 29, 2017).
9. Crawford K, Miltner K, Gray ML. Special section introduction. *Int J Commun*. 2014;8:1663–1672.
10. Anderson C. The end of theory: The data deluge makes the scientific method obsolete. *Wired* 2008, June 23.
11. Mayer-Schönberger V, Cukier K. *Big data: A revolution that will transform how we live, work and think*. London: John Murray, 2013.
12. Pentland A. *Social physics: How good ideas spread-the lessons from a new science*. New York: Penguin, 2014.
13. Vertesi J, Pappalardo R, Alexander C, et al. Sociological considerations for the success of planetary exploration missions: A white paper submitted for consideration to the Planetary Science Decadal Survey 2013–2022 Committee. 2009:1–8. Available at: [http://www8.nationalacademies.org/ssb/survey/DetailFileDisplay.aspx?id=199&parm\\_type=PSDS](http://www8.nationalacademies.org/ssb/survey/DetailFileDisplay.aspx?id=199&parm_type=PSDS) (accessed June 7, 2017).
14. Viseu A. Integration of social science into research is crucial. *Nature*. 2015;525:291.
15. Fiore-Gartland B, Tanweer A. Community-level data science and its spheres of influence: Beyond novelty. *eScience Institute Blog*, 2015. Available online at <http://escience.washington.edu/community-level-data-science-and-its-spheres-of-influence-beyond-novelty-squared/> (last accessed March 30, 2017).
16. Drouhard M, Tanweer A, Fiore-Gartland B. A typology of hackathon events. In: *Hacking at Time-Bound Events Workshop at Computer Supported Cooperative Work 2016 (CSCW'16)*, 2016, pp. 1–4.
17. Tanweer A, Fiore-Gartland B, Aragon C. Impediment to insight to innovation: Understanding data assemblages through the breakdown–repair process. *Inf Commun Soc*. 2016;19:736–752.
18. Dossick CS, Neff G, Osburn L, et al. Technical boundary spanners and translation: A study of energy modeling for high performance hospitals. *Engineering Project Organization Conference*, Cle Elum, WA, 2016.
19. Monson C, Dossick CS, Neff G, et al. Finding connections between design processes and institutional forces on integrated AEC teams for high performance energy design. *Engineering Project Organization Conference*, Cle Elum, WA, 2016.
20. Gitelman L, Jackson V. Introduction. In: Gitelman, L. (Ed.): “Raw Data” is an oxymoron, Cambridge, MA: MIT Press, 2013, pp. 1–14.
21. Bowker GC. Memory practices in the sciences. Cambridge, MA: MIT, 2005.
22. Martin A, Lynch M. Counting things and people: The practices and politics of counting. *Soc Probl*. 2009;56:243–266.
23. Fiore-Gartland B, Neff G. Communication, mediation, and the expectations of data: Data valences across health and wellness communities. *Int J Commun*. 2015;9:1466–1484.
24. Pasquetto IV, Randles BM, Borgman CL. On the reuse of scientific data. *Data Sci J*. 2017;16:1–9.
25. Hargittai E. Is bigger always better? Potential biases of big data derived from social network sites. *Ann Am Acad Pol Soc Sci*. 2015;659:63–76.
26. Tufekci Z. *Engineering the public: Big data, surveillance and computational politics*. First Monday 2014;19(7).
27. Dourish P. What we talk about when we talk about context. *Pers Ubiquitous Comput*. 2004;8:19–30.
28. Seaver N. The nice thing about context is that everyone has it. *Media Cult Soc*. 2015;37:1101–1109.
29. Fiore-Gartland B, Tanweer A. It's the context, stupid: Reproducibility as a scientific communication problem. *4S/EASST Conference*, Barcelona, 2016.
30. Borgman CL. The conundrum of sharing research data. *J Am Soc Inf Sci Technol*. 2012;63:1059–1078.
31. Borgman CL. *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press, 2015.
32. Downey GJ. Making media work: Time, space, identity, and labor in the analysis of information and communication infrastructures. In: Gillespie T, Boczkowski PJ, Foot KA (Eds.): *Media Technologies*, Cambridge: MIT Press, 2013, pp. 141–165.
33. Edwards PN. *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge, MA: MIT Press, 2010.
34. Edwards PN, Jackson SJ, Chalmers MK, et al. *Knowledge infrastructures: Intellectual frameworks and research challenges*. Ann Arbor: Deep Blue, 2013.
35. Ribes B, Jackson S, Geiger S, et al. Artifacts that organize: Delegation in the distributed organization. *Inf Organ*. 2012;23:1–14.
36. Ribes D, Jackson S. *Data Bite Man: The work of sustaining long-term data collection*. In: Gitelman, L. (Ed.): “Raw Data” is an oxymoron, Cambridge, MA: MIT Press, 2013, pp. 147–166.
37. Star SL, Strauss A. Layers of silence: The ecology of visible and invisible work. *Comput Support Coop Work*. 1999;8:9–30.
38. Vertesi J. *Seeing like a Rover: How robots, teams, and images craft knowledge of Mars*. Chicago, IL: University of Chicago Press, 2015.
39. Vertesi J, Dourish P. The value of data: Considering the context of production in data economies. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW'11)*, 2011, pp. 533–542.
40. Kitchin R. *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage, 2014.
41. Abolafia MY. Narrative construction as sensemaking: How a central bank thinks. *Organ Stud*. 2010;31:349–367.
42. Sharon T, Zandbergen D. From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media Soc*. 2016;1–15.
43. Levy KEC. Relational big data. *Stanford Law Rev Online*. 2013;66:73.
44. Couldry N, Powell A. Big Data from the bottom up. *Big Data Soc*. 2014;73:79.
45. Gregory J, Bowker GC. The data citizen, the quantified self, and personal genomics. In: Nafus, D. (Ed.): *Quantified: Biosensing technologies in everyday life*, Cambridge, MA: MIT Press, 2016, pp. 211–226.
46. Hautea S, Dasgupta S, Hill BM. Youth perspectives on critical data literacies. *ACM Conference on Human Factors in Computing Systems (CHI 2017)*, Denver, CO, 2017, pp. 919–930.
47. Aragon CR, Bayer J, Echenique A, et al. Developing a research agenda for human-centered data science. *Human-centered Data Science Workshop, CSCW 2016: ACM Conference on Computer Supported Cooperative Work*, San Francisco, 2016.
48. boyd d. Where do we find ethics? *Points*, 2016. Available online at <https://points.datasociety.net/where-do-we-find-ethics-d0b9e8a7f4e6> (last accessed March 29, 2017).
49. Schrock A. What communication can contribute to data studies: Three lenses on communication and data. *Int J Commun*. 2017;11:701–709.
50. Neff G, Nafus D. *Self-tracking*. Cambridge, MA: MIT Press, 2016.
51. Neff G. Why big data won't cure us. *Big Data* 2013;1:117–123.
52. Kuksenok K. Influence apart from adoption: How interaction between programming and scientific practices shapes modes of inquiry in four oceanography teams. PhD dissertation, Department of Computer Science and Engineering, University of Washington, Seattle, WA, 2016.

**Cite this article as:** Neff G, Tanweer A, Fiore-Gartland B, Osburn L (2017) Critique and contribute: a practice-based framework for improving critical data studies and data science. *Big Data* 5:2, 85–97, DOI: 10.1089/big.2016.0050.

#### Abbreviations Used

EUI = energy use index  
OSM = OpenStreetMap