

Feminist Data Studies

MARY ELIZABETH LUKA

University of Toronto Scarborough, Canada

KOEN LEURS

Utrecht University, Netherlands

Introduction

Data, datafication, and data studies have become key buzzwords characterizing contemporary life in the digital and knowledge economies. Critiques and ways to mobilize data for social and economic benefit are variously—and often contradictorily—imagined by scholars in disparate disciplines, as well as by politicians and policy makers, corporations and nonprofit organizations, citizens, and activists. The datalogical turn to study the structuring of data foregrounds the proliferation of algorithmic processing and data as an emergent regime of power and knowledge and value exchanges in the digital datafication of everyday life and culture. Big Data, digital methods, and data studies are more than simply semantic currencies neutrally describing present-day conditions. Indicative of Big Data as a privileged mode of knowledge production, all too often, quantitative, abstracted, and disembodied approaches are privileged over qualitative data approaches. However, we argue, alongside many others, that database technologies and human experiences are always necessarily mutually constituted (Metcalf & Crawford, 2016). Infrastructures, categorizations, and algorithmic processing are commonly black-boxed and therefore invisible with the consequence that data generated is never raw, but always cooked (Bowker, 2006). Data, data analysis, and data visualizations are never neutral, but are power ridden, situated, as they are subject to choices made by humans and machines. Yes, we “are” data, as evidenced by the way that a growing number of the world’s populations are increasingly rendered as datafiable (see, for example, Cheney–Lippold, 2017). Data typologies are inescapably based on a moral agenda that prioritizes one worldview over many others. The military-industrial data analysis machinery reestablishes boundaries (and therefore barriers) between “data-haves” and “data have-nots,” which are commonly based on categories of difference. Processes of datafication of culture and populations are never devoid of the various forms of cultural prejudices and discriminations. Rather, they are often used to exacerbate intersectional power hierarchies based on gender, sexuality, race, class, ability, religion, migration status, and age.

In this entry, we draw from the principles of feminist ethics of care—which include attention to human meaning-making, situatedness, context-specificity, dependencies, and relationalities—to elaborate what feminist data studies could look like (Leurs, 2017;

The International Encyclopedia of Gender, Media, and Communication. Karen Ross (Editor-in-Chief), Ingrid Bachmann, Valentina Cardo, Sujata Moorti, and Marco Scarcelli (Associate Editors).

© 2020 John Wiley & Sons, Inc. Published 2020 by John Wiley & Sons, Inc.

DOI: 10.1002/9781119429128.iegmc062

Luka & Millette, 2018). As part of an international, loosely networked collaborative working group, we also draw on the work of our colleagues and discussions within the group, and well outside it. The group of 13 multi- and interdisciplinary scholars includes the authors of this entry and Rena Bivens (Carleton), M el Hogan (University of Calgary), Alison Harvey (Leicester), Jessalyn Keller (University of Calgary), M elanie Millette (UQAM), Sarah Roberts (UCLA), Tamara Shepherd (University of Calgary), Jacqueline Wallace (Concordia), and Andrea Zeffiro (McMaster). As feminists committed to enacting intersectional (Crenshaw, 1989) and inclusive approaches to analysis, we aim to situate lesser-known communities of scholarship as new centers or zones of enquiry in the multidisciplinary fields of research within which data studies takes place. Moving from critique to generative approaches, this entry first presents an overview of critical takes on data studies, including examples of more inclusive methods and approaches. In the last section, we present ways to engage in data studies, rooted in social justice in the social sciences and humanities.

Critique

Dictionary definitions of the term “data” encapsulate the challenge of unpacking deeply embedded and often contradictory understandings of the term itself. Such simplified definitions understand data as a series of “facts,” or as (somehow neutral) zeros and ones generated by computers. This approach de-emphasizes the manufactured and specific nature of how data is constantly resituated in time, place, and context.

Internet and data scholars suggest that the quantifiable concept of data opposes the processual conditions (and depth of knowledge) that is available from many forms and types of data, and also that this approach is deeply embedded in specific commercial and decision-making processes. Indeed, scholars across disciplines agree that any data collection is the result of a process of decisions, whether those are made by computer programmers and builders, computer analysts, and platform stakeholders (Bowker, 2006; Parks & Starosielski, 2015), or by qualitative approaches that are sensitive to the conditional nature of data (Leurs, 2017). In the digital era, researchers (scholarly, open access, commercial or otherwise) work hard to find innovative ways to use software and hardware—as well as more traditional, qualitative approaches—to identify, scrape, archive, visualize, and make sense of data. Data, however, is shaped by how it is collected, curated, and sorted in ways that may seem neutral but are not. Borgman (2016) insists on the open potential of data and asks us to contextualize data, including when and how it is created, collected, and used, as well as how those contexts and meanings change over the course of its existence as data. This not only signals the temporal and shifting nature of data itself, but also our evolving understandings of data, its expression, and its analyses.

To help pinpoint the emergent and iterative nature of feminist data studies, we use the definitions of “big,” “small,” “thick,” and “lively” data delineated by Luka and Millette (2018) as a foundation for building toward a provisional understanding of feminist data studies. Like Borgman, they note that data categories and registers emerge

from research questions and objectives, as well as from methodological approaches. These categories and registers can be complementary though also sometimes stand in contradiction to other datasets. Consequently, they also note that quantitative (often computerized) strategies are not always sufficient for understanding the hows and whys of the world that we seek to analyze. Of course, such strategies can be usefully engaged in the processing of “big” data, whose characteristics include high volume, breadth, the ability to be coded and indexed electronically, and, often, to illustrate relationality within and among data sets. Complementing automated Big Data processing activities, however, other forms of qualitative data can be processed in more human-centered ways, often as “thick” or “small” data. Thick data derives from ethnographic methods and is often mobilized to analyze complex and sometimes site-specific or informal relationships such as emotions, worldviews, or identities. Often involving detailed single or multiple case studies or participant observation, the accretion of layers of data generated through immersion provide in-depth understandings of contexts, participant experiences and knowledge, and varied levels of information to process over time. Not incompatibly, small data is characterized by its scale. Luka and Millette (2018) take up well-established definitions of small data as analyzable by hand, and by the comprehensibility of such data without computer assistance. In contemporary complex data studies, as Borgman suggests (above), it is also important to acknowledge the temporal nature of everyday data. Luka and Millette (2018) augment Deborah Lupton’s concept of “lively” data, which arises from daily online interactions and intersections, and includes the computerized processes used to produce that data, to note that such data operates at both meso- and microlevels, but also over time, and through different data infrastructures and data centers.

From a feminist perspective, it is important to acknowledge the materiality of the collection and distribution systems engaged in collecting, sorting, and analyzing data within all of these registers on an ongoing basis. This includes, for example, the sometimes environmentally damaging and territorially ambiguous volume of work that takes place in data centers or the smaller but equally environmentally fraught nature of taking multiple airplane trips to a data collection site requiring long-term, in-depth exposure. The comprehensive anthology on the materiality of data centers and platforms developed by Parks and Starosielski (2015) notes that any investigation of datafied material must also include the often invisible but still visceral nature of computerization and data analysis, including servers, human coders and the way that information is processed in batches or packets.

What this configuration of modes of data reveal are pathways to question the dominance of White, Western-centric, middle-class, able-bodied, and heteronormative approaches to the generation and analysis of data. Resistance to such hegemonic practices are realized by making platform ownership visible; by revealing the amplification of discrimination, violence, and hate through user norms and content moderation practices; and by calling digital data researchers to account for visualization and related practices that fetishize data or claim a false neutrality.

Infrastructure and Datafication

The workings of infrastructures that we rely on every day to generate, access, store, circulate, and share data include cloud computing and data storage, transatlantic internet backbones, electricity grids, Wi-Fi, 4G connectivity, as well as social media platforms that often remain totally invisible to us. Feminist and postcolonial technoscience scholars, such as that evidenced by Sandra Harding's decades of work on understanding how women and other marginalized groups have been sidelined in the sciences both conceptually and practically (e.g., Harding & Hintikka, 2003), have developed generative conceptual and methodological tools to understand the inner workings of infrastructures. They draw attention to its commodified forms, including the paradox generated by aiming to create an open data movement from and for publicly funded or generated research and information that must be restricted or is incomprehensible. Taking a relational approach to societies and technologies, this strand of scholarship draws attention to how these two mutually shape and co-construct each other. For example, Donna Haraway (1997) describes this approach as a way to understand the aggregative and ever-changing nature of relations between humans and nonhumans. Most poignantly, these experts call into question the power relations that are baked into infrastructures, asking why infrastructures materialize and reinforce dominant worldviews and infrastructures that replicate themselves rather than being open to multiple points of view or pathways. In this sense, understanding the potential multiple architectures of (or lenses, or perspectives on) data is an important way to scrutinize power hierarchies and decision-making that lead to inclusion or omission, and prioritizing and silencing, during every moment of relational assemblage. Infrastructures are not dormant, but demand constant actualization, reinforcement, and input.

This is not simply a theoretical speculation (though there is plenty of useful work to be done by mobilizing feminist materialist speculation—see “Generative” below). First, the immateriality and invisibility of data related infrastructures demands our attention. In order for there to be data, precarious manual labor is performed to mine minerals and produce consumer goods. In his book, *Goodbye iSlave* (2016), Jack Qui has documented parallels between assembly line workers in the Foxconn factories in China producing smart phones and the transnational slave trade and indentured labor during the colonial era. He, for example, points out the macabre similarities between the safety nets applied around high-rise Foxconn factory workers dormitories that prevent suicide attempts to similar nettings used on slave ships to prevent people from jumping overboard, choosing death over slavery. Furthermore, when devices and data infrastructures are used on increasingly exponential levels, they lead to e-waste and exhaust our environment, as Mél Hogan (2015) points out. Focusing on the case of a data center in the drought State of Utah—data centers depend on massive water resources to offer its services—Hogan raises important questions about the ownership and depletion of natural resources including water and land.

Moreover, datafication seeks to make human subjects legible in sociotechnical systems, and the infrastructural moment of reading bodies often automatically includes selection. Selecting some subjects implies discriminating against others. As Safiya Noble

argues in *Algorithms of Oppression* (2018), search engines reinforce racism: in its presentation of findings, human subjects are typically hierarchically ordered in a limited, binary way as desirable and undesirable.

In combination with this effect, and alongside the black-boxed workings of datafication, attention is needed to parse platform capitalism's impact upon datafication. Platforms structure entire economies (not just companies): that is, corporate owners of platforms commodify user-generated data (user activities, postings, photos, videos, etc.), but also maintain uneven access to data (corporate analytics prioritized over academic research) and maintains the ability to alter restrictions as it sees fit. Within the global academic landscape, the proliferation of academic Big Data research is accelerating the uneven politics of knowledge production that reinforces divisions between scholars residing in the Global North with larger research budgets that could support paying for data access and those residing in the Global South working with smaller budgets. Feminist data studies interventions can draw on understandings of the underpinnings of infrastructures as sociotechnical assemblages which are not neutral or innocent, but reflective of and constitutive of unjust broader societal configurations. Feminist data studies helps point out how data can harm people, and how it can have strong impact on the future of individuals, particularly vulnerable groups such as refugees. For example, in Europe, asylum seekers from the Middle East and sub-Saharan Africa risk being deported based on their digital records, when on arrival to Europe they are asked to provide fingerprints and other biometrical data which are algorithmically processed in search of data-doubles in the European Dactyloscopy Database (EURODAC) (Leurs & Shepherd, 2017). Moreover, in the Netherlands, the government has embraced System Risk Indication (SyRi) to distinguish between "nonrisk" and "high-risk" citizens. While this approach breaches the data protection principle of purposeful limitations (as individuals are not informed their data is processed in connected databases), the resulting questionable evaluations of human propensity to violate social norms and international rules (i.e., to behave in increasingly risky ways) are also based on analyses of data collected for other, unrelated purposes, that is, on the basis of cross-referencing taxpayer information with health care and migration status data (Oosterloo & Van Schie, 2018). Such cross-referencing has not been proven to predict risky behavior.

Research Norms/Hierarchies Online/User Norms

Gender, class, and identity investigations from digital gender, media, and production studies provide helpful cues about how power dynamics influence not only what gets produced as data, but also who, what, how, and when data is circulated as content and power. Such analysis is useful, most particularly, through the distribution of some findings or the promulgation of some contexts, but not others. In other words, we need to ask "who is data, as well as when and how and why; and who is asking?" Although scholars in these fields have long and repeatedly revealed production and distribution systems as gendered and racialized, in a datafied environment, we must also unpack how transnational media cultures are shaped by the circulation of meaning

(as data) and forms of distribution as the production of discriminatory and powerful infrastructures.

Moreover, indigenous scholars and scholars of indigenous ways of thinking in related, broader fields of enquiry such as science, philosophy, and the arts make powerful arguments to give credit where credit is due, including to reveal and correct silences, elisions, and appropriations. For example, traditional knowledge, or traditional ecological knowledge, in the sciences, social sciences, and humanities (e.g., Todd, 2015) has long pointed to the way forward for feminist data studies to become allied with the compelling thickness and liveliness of how site- and culturally specific knowledge can articulate meaning around the land-as-data, including multispecies and spiritual considerations. Complexity is required here too. Some feminist and indigenous data studies assert the right of indigenous people to collect and not share data in nation-to-nation relationships (colonial + indigenous). To better understand this position, indigenous data sovereignty is explicated through open projects such as <http://usindigenousdata.arizona.edu/>. Debates about data ownership and deployment are not, of course, limited to indigenous initiatives. An international debate erupts with great regularity in the global sphere, including at UNESCO, around identity and cultural rights articulated as matters of cultural sovereignty, privacy, and surveillance.

In a 2018 special issue of *Feminist Media Studies* on online misogyny in data studies, the issue as a whole demonstrates that Big Data (including open access and crowd-sourced projects) and digital humanities research continue to be characterized by racial inequalities. They are not the first to do so. In 2012, Nakamura and Chow-White published a collection of essays in *Race After the Internet* that helped lay the groundwork for the development of feminist data studies. The essays mobilized an intersectional approach to data analysis from ethnography, social science, text, media, and rhetorical fields. Gathering up studies that consider data generated in everyday activities (e.g., social media, video games, online videos, health, and education), *Race After the Internet* delineated how racial profiling and surveillance permeate contemporary coding, interaction, and analysis. Many of the contributors, including danah boyd, Kate Crawford, and Tara McPherson have since developed additional research about how datafied social media environments reinforce racism and power dynamics, amplifying these conditions as entrenched, incompatible, and unchangeable. As the field develops, not only is the relational or co-constitutive nature of data collection and analysis important to understand, then, but more explicitly, so is the way in which racism, class, sexism, and colonialism is built in to the academy's privileging of quantitative data-based research (Chakravartty, Kuo, Grubbs, & McIlwain, 2018). For example, Chakravartty et al. report on their large-scale study of authorship and citation practices in communication studies, and demonstrate statistically significant patterns of segregation and disproportionate representation of White men in citation practices and editorial boards. This is also a core thematic of recent analyses of the infrastructure and data collection and processing practices of the digital humanities, including at significant research and discussion websites such as <http://dhdebates.gc.cuny.edu/>. Not surprisingly, the uneven politics of Big Data

knowledge production reinforces divisions between the Global North and the Global South.

Moreover, activities described as scraping, mining, or other masculine forms of conquest not only continue to reinforce such divisions, but also to present a lucrative frontier for science and technology studies. In some of the most recent scholarship on this issue, Mascarenhas (2018) argues that racism and the specificity of race in relation to science, technology, and society (STS) has been ignored, including how racism continues to fundamentally shape contemporary sociocultural and political conditions. This is a key component of the so-called “black box” of STS. Similarly, Borgman (2016) argues that to unpack the black box of STS, and data studies in general, we must understand size, timing, the nature of what “open access” means in different environments, how knowledge is understood and articulated *and* the economic context (commercializability, property strictures, and intellectual property rights), as well as the contextual ethics of each discipline. She does not, however, unpack how specific participants, subjects, and communities ought to be analyzed or empowered, pointing to a significant gap in understanding the power relationships embedded in several characteristics of STS and data studies that still require parsing. In contrast, Anable (2018) notes that the legacy of discrimination revealed in feminist media studies allows us to more carefully examine feminist platform studies, including how such platforms are developed, built, used, and impact on already existing media and mediated infrastructures. Likewise, feminist data studies calls into question emerging uses of artificial intelligence. For example, facial recognition techniques have been programmed to determine whether people are homosexual or lesbian on the basis of their facial features (Lewis, 2018), marking some sexual preferences as undesirable, thereby perpetuating inequalities.

Visuality

Perhaps one of the most egregious examples of the impact of such omissions are found in the deployment of visuality throughout the field. Several contemporary social media platforms are primarily visual, or employ a combination of visuals and text, and little work has been done to date to understand how these elements involve or impact users from methodological and ethical perspectives. There are several feminist scholars who have long critiqued how data visualizations construct conventions of objectivity, validity, and neutrality, based on appeals to ideologies of “transparency, scientific-ness and facticity” (Kennedy, Hill, Aiello, & Allen, 2016, p. 716). Such mappings reflect colonial forms of visuality as elite distant observers decide who is mapped, for what purposes, on the basis of their top-down directed parameters. For data-based visualization projects such as Lev Manovich’s participatory Instagram project of Visual Earth (<http://manovich.net/index.php/exhibitions/visual-earth>), much more attention could be paid to how aggregations of the visual obfuscates the deeply classed, racialized, and gendered nature of this form of (mediated) production. While deliciously attractive as visualizations, the starbursts and other representations also obscure the overwhelming accretion of images circulated by only a narrow segment of society—those with access to digital devices, software apps, and the internet.

Moreover, while satellite imagery spans the entire globe, there are serious gaps in the location-based information that Google Maps provides. Various substantial urban and rural areas in the Global South are not mapped. As the artist Mimi Onuoha captures in her installations and writings on algorithmic violence, in Rio de Janeiro, only 2% of the favelas are mapped, leaving over one million people, or a quarter of the city's inhabitants, off the map. The same holds true for the 100,000 inhabitants of Makoko, a floating lake community off the coast of Lagos, Nigeria, and the almost 2 million inhabitants of Chad's Lake Fitri drainage basin (see <http://mimionuoha.com>). Primarily used for marketing, monetization, exploitation, and surveillance, deploying data through visualization is an exemplar for the extractive nature of some Big Data exercises, reminiscent of military-industrial and colonizing strategies of many centuries past.

Another case in point is the politics of data visualization employed by the International Organization for Migration. It provides an accessible-for-all dashboard that tracks missing migrants along frequent migratory pathways. As of October 15, 2018, the website claims, "only" 2,902 migrant fatalities were registered, when you select the option "Europe" from the drop-down menu (<https://missingmigrants.iom.int>). However, this low number does not acknowledge that Europe is currently the deadliest migration destiny in the world. Deaths at sea are strategically excluded from the European visualization. It is only by clicking "Mediterranean" from the drop-down menu that we can see what is happening at the borders of Europe. The territory of Europe here is falsely demarcated from the Mediterranean sea mass. While these visualizations based on aggregated data sets are presented as accurate and "neutral" representations, they are subject to the political choices and power-ridden situated context of production. Such misleading examples of data-harm are tracked on websites such as those operated by the Data Justice Lab at Cardiff University in the United Kingdom (<https://datajusticelab.org/data-harm-record>).

As STS visualizations multiply, it is helpful to know that tools to illustrate multi-perspectival approaches are beginning to emerge. Such approaches demonstrate how to interrupt or disrupt the overpowering normative structuring within most visualization software, pointing us toward the deeper activation of feminist data studies through visualizations.

Generative: Some Practical Considerations

There is quite a lot of emergent ground in data studies, with scholars developing research that incorporates critical perspectives to counteract the replication of previous power dynamics and oppressions. Often, a feminist ethics of care in a digital setting can draw attention to the power-ridden, relational, and context-specific dynamics of knowledge production, raising questions about how we may care for human subjects and take the time to acknowledge our situatedness in data research. We take inspiration from criteria of anti-oppressive research developed by Bhavnani (1993): how data can be used while avoiding the "reinscription" of power hierarchies, how we can account for the "micropolitics" of data research praxis, and how we can foreground the integrity of "difference."

Åsberg, Thiele, and Van der Tuin (2015) build on the idea of situated knowledges to enact a feminist materialist speculative methodology. Activating research *with* others rather than conducting it *on* others, is a material act of performative speculation which enables rather than masks multiple subjectivities and asserts the importance of acknowledging positionality, contexts, and biases as a key source of rich, in-depth knowledge generation. To activate commitments to equity, diversity, and social justice at the individual as well as collective levels, Åsberg et al. (2015) ask researchers to imagine how contextual circumstances articulate appropriate situated knowledges, and how we can unpack deeply embedded binary categories of data in many social sciences, arts, and humanities traditions. It is not just that intersectional feminist practices of care contest findings that divide the world into typologies based on absolutist binaries. It is also that we must move beyond binaries, including the division of nature and culture. Indeed, we need to pay attention to all registers of life and the concomitant power dynamics in this period we call the Anthropocene, that is, the era that is most affected by human intervention. So how can we go about this work? A significant body of literature has developed which addresses exactly this question.

Practicalities

Luka and Millette (2018) close out their argument for a feminist ethics of care with a series of practical questions to ask during research design, collection, and analysis phases, drawing on participatory action research and research-creation experiences of their own and others. They also draw on evolving professional standards, including those under revision at the international Association of internet Researchers. These include understanding positionality and developing shared power dynamics as well as the potential for collaboration and making data collection visible, iterative, and debatable. It also includes understanding why and whether collaboration or simply more shared involvement makes sense, and explicitly asks how the work will help build equity not just in human relations, but in the relationship of humans to the rest of the world. They pointedly ask who funds research, who benefits—and who suffers—from it. Feminist activists have also collaboratively published a collection of principles for research and action on the internet (www.feministinternet.org) ranging from addressing who has access, to issues around open-source, memory, violence, resistances, and movement-building.

To other scholars, such as Massanari (2018), increased visibility also increases the risk to feminist data researchers, by becoming subject to the gaze (and sometimes the social mediated ire) of far-right groups. A recent proliferation of guidelines for safe internet practices includes the DIY Guide to Feminist Cybersecurity (<https://hackblossom.org/cybersecurity>), and blog posts about how to protect yourself as a researcher while conducting research about groups opposed to equity, or simply in favor of the status quo.

While it is an oxymoron to think about data as raw (Bowker, 2006), since data are never a “natural” rendering of reality, there are clearly levels, registers, actors and participants, contexts, and intentions that may be mappable from a variety of perspectives.

Although data sets can be massive, they are always limited representations of reality. Even when these are context and time-stamped, the coding and mapping of the data and consequent findings are too often rendered ahistorical and decontextualized.

When conducting social media research on Twitter discussions on #feminism, Kirkwood, Cree, Winterstein, Nuttgens, and Sneddon (2018) lament that as third party researchers they could not obtain access to all tweets but could only gather a small proportion of data via the API. In such cases, researching the stories behind the data can provide critical insights as well. For example, reflecting on the Quipo Project (quipu-project.com) an interactive digital counter-memory archive of unconsented sterilization in Peru, researchers Brown and Tucker (2017) argue that relationships with participants must be carefully developed to enable trust and shared understandings. This approach is consistent with long-established feminist research approaches. With such a foundation, it becomes possible to build a rich data set that can inform larger contexts and similar situations.

The mechanisms through which data sets are processed and analyzed can be both demystified and carefully balanced through a social justice orientation. Therefore, alongside questions such as “when are data,” and “when is an architecture,” during the phase of data analysis, it is important to reflect on the question “how am I processing the data that I am collecting?” Analysis commonly starts with the phase of data sorting and data cleaning. As extractive corporate platforms have not structured their data with social justice-oriented researchers in mind, data sorting can pose serious questions. Consider, for example, the long history of Facebook and its inability to collect and manage more than binary gender identifications of users, charted by Rena Bivens (2017). In response to concerns shared by LGBTQ2S activists, the platform included a range of nonbinary sexual identification in its profiling categories. By looking more deeply into the backend of the coding system, however, these profiling options were found to be subsumed to data preferences and choices made by Facebook participants that thereby assigned their gender according to the male/female gender binary (Bivens, 2017). Such elisions in data collection practice and processing are forms of algorithmic violence that feminist data studies seeks to reveal and to propose alternative approaches.

Besides data sorting, data analysis also involves cleaning, selection, and coding procedures which are not commonly well documented. A notable exception is Kirkwood et al. (2018), who offer reflection on their difficulties of developing their artificial intelligence (AI) classifier to auto-tag tweets dealing with #feminism. After rounds of manual tagging to train their AI classifier, they found that additional manual tagging was needed to improve their machine learning algorithm. Emergent research also demonstrates that large-scale data patterns can be made particularly meaningful when brought into interaction with small scale approaches that embrace ambiguity and complex meaning making. For example, Concannon, Balaam, Simpson, and Comber (2018) conducted a feminist geography study on a large-scale dataset of reviews about breastfeeding in public spaces, and they propose a reflexive process that combines large-scale aggregation of government and user-generated data with a qualitative engagement by mobilizing user-generated data that can challenge and recontextualize official data.

Data visualizations are increasingly mainstreamed tools, and discussions continue to unfold about feminist data visualization strategies. Mahmud, Hogan, Zeffiro, and Hemphill (2017) propose that in teaching students data visualization, students can be trained to find ways to use information visualization that reveal rather than conceal and that unpack possibilities rather than synthesize or exclude information. Mahmud et al. further suggest students can do so by providing context and comparisons to illuminate the research object as fulsomely as possible, which will then enable participants to be connected to potential audiences or researchers who wish to study insights based on the research at hand. Similarly D'Ignazio (2017) proposes that feminist data visualizers need to reimagine gender binaries and take up intersectional and pluralistic approaches. D'Ignazio further suggests that one of the ways that empowerment among marginalized groups can be realized is not only by better understanding power dynamics but also by understanding how labor is embodied in such groups in visual ways. Academics have mobilized alternative data studies to document injustices, exploitation, and marginalization (Chakravartty et al., 2018). The #seperados #tornapart postcolonial digital humanities project conducted by Columbia's Group for Experimental Methods in the Humanities in collaboration with Borderlands Archives Cartography is a strong example of how social justice-oriented data visualization can be mobilized in feminist data studies. Torn Apart is a visualization database that uses xPMethod to operate in real time (<http://xpmethod.plaintext.in/torn-apart/index.html>), tracking the impacts of the United States' 2018 Zero Tolerance Policy on asylum seekers who seek entry into the United States.

When conducting large-scale Big Data research on Twitter, interdisciplinary teams are needed, consisting of colleagues trained in computer science and artificial intelligence, as well as those versed in critical theory and feminist toolboxes. Things do not always go easily, however. In order to conduct and assess interdisciplinary work that can produce broad as well as deep understandings of data sets, researchers often struggle to find points of connection and comparison. For example, Kirkwood et al. (2018) reveal that during their large-scale Twitter study, their mutually agreed undertakings to realize feminist values throughout the work they were doing—including a shared commitment to challenge gender discrimination—were not based on the same assumptions or terminology in their respective fields. Those who have attempted to work across disciplinary fields will recognize these difficulties, while those familiar with the challenges that come from intersectional and gender discrimination will understand how difficult managing power dynamics within research teams or with participants can become, no matter how well-intentioned.

A Final Word

Over the last decade, data studies scholarship has developed in ways that increasingly take account of the capacity of seemingly “soft” feminist social science methodology to develop broad and comprehensive examples of quantitative work related to finding ways to understand daily life in all its complexities, including qualitative insights that could be drawn from quantitative patterns. For example, Scott and Siltanen (2017) have rethought regression analysis as a feminist undertaking by finding ways to analyze three

elements of an intersectional approach, including context as well as the heuristics and the intersecting complexities involved in systemic inequality. Their cues about how to collect, code and correct for traditional quantitative social science methodologies provide a useful pathway for feminist data studies to explore. And finally, Bucher (2018) and Noble (2018), among others, offer in-depth criticisms of software that thoroughly unpack the incredible power dynamics that underpin the mediated worlds of algorithms. Much more work remains to be done in this arena, however, as emerging scholarship continues to demonstrate.

In all, feminist data studies seek to mobilize data-driven insights with a commitment to social justice, and with an awareness that data, datafication, and data visualizations are shaped not just by the intentions of people involved, but also by the possibilities and limitations of computational systems. Data frequently emerge from, and are pre-constructed within, a context of corporate extraction, environmental destruction and exploitation, precarious labor, weak regulation, and limited accountability. Feminist ethics of care offers tools to move beyond pursuing value-free universal truth claims and disembodied generalizations but self-reflexive, situated, and partial accounts.

SEE ALSO: Apps, Sex, and Quantified Self; Dating Apps; Digital Divide; Digital Labor; Feminism, Media, and the Public Sphere; Feminist/Activist Responses to Online Abuse; Feminist Press; Gamergate; Gender and Media; Gender, Media, and Political Economy; Gender and Technology; Gendered Hate Online; Gendered Identities Online; Globalization, Gender, and Media; LGBTQI Online; Newsroom Cultures; Online Identities and Gender Norms; Postfeminist Media Cultures; Social Media: Complexities and Contradictions; Trans Cultures Online; Visual Representations and the Feminist Challenge; Waves of Feminism

References

- Anable, A. (2018). Platform studies. *Feminist Media Histories*, 4(2), 135–140. doi:10.1525/fmh.2018.4.2.135
- Åsberg, C., Thiele, K., & Van der Tuin, I. (2015). Speculative before the turn: Reintroducing feminist materialist performativity. *Cultural Studies Review*, 21(2), 145–172. doi:10.5130/csr.v21i2.4324
- Bhavnani, K. K. (1993). Tracing the contours: Feminist research and feminist objectivity. *Women's Studies International Forum*, 16(2), 95–104.
- Bivens, R. (2017). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society*, 19(6), 880–898.
- Borgman, C. L. (2016). *Big Data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Bowker, G. C. (2006). *Memory practices in the sciences*. Cambridge, MA: MIT Press.
- Brown, M., & Tucker, J. (2017). Unconsented sterilization, participatory story-telling, and digital counter-memory in Peru. *Antipode*, 49(5), 1186–1203. doi:10.1111/anti.12316
- Bucher, T. (2018). *If ... then: Algorithmic power and politics*. Oxford, UK: Oxford University Press.
- Chakravartty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). #CommunicationSoWhite. *Journal of Communication*, 68(2), 254–266. doi:10.1093/joc/jqy003

- Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York: NYU Press.
- Concannon, S. J., Balaam, M., Simpson, E., & Comber, R. (2018). *Applying computational analysis to textual data from the wild: A feminist perspective*. *Proceedings of CHI 2018*, April 21–26 (paper 226). Montreal, Canada: ACM. doi:10.1145/3173574.3173800
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), Article 8. Retrieved from <http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>
- D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), 6–18. doi:10.1075/idj.23.1.03dig
- Haraway, D. (1997). *Modest_witness@second_millennium: Femaleman_meets_oncomouse: Feminism and technoscience*. New York, NY: Routledge.
- Harding, S., & Hintikka, M. B. (Eds.). (2003). *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and philosophy of science*. Dordrecht, Netherlands: Kluwer Academic.
- Hogan, M. (2015). Data flows and water woes: The Utah data center. *Big Data & Society*, 2(2). doi:10.1177/2053951715592429
- Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualization conventions do. *Information, Communication & Society*, 19(6), 715–735.
- Kirkwood, S., Cree, V., Winterstein, D., Nuttgens, A., & Sneddon, J. (2018). Encountering #feminism on Twitter: Reflections on a research collaboration between social scientists and computer scientists. *Sociological Research Online*, 23(4), 763–779. doi:10.1177/1360780418781615
- Leurs, K. (2017). Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Review*, 115(1), 130–154. doi:10.1057/s41305-017-0043-1
- Leurs, K., & Shepherd, T. (2017). Datafication and discrimination. In K. van Es & M. Schäfer (Eds.), *The datafied society: Studying culture through data* (pp. 211–234). Amsterdam, Netherlands: Amsterdam University Press.
- Lewis, P. (2018, July 7). “I was shocked it was so easy”: Meet the professor who says facial recognition can tell if you’re gay. *The Guardian*. <https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis>
- Luka, M. E., & Millette, M. (2018). Re(framing) Big Data: Activating situated knowledge and a feminist ethics of care in social media research. *Ethics as method* (Special issue). *Social Media & Society*, 4(2).
- Mahmud, A., Hogan, M., Zeffiro, A., & Hemphill, L. (2017). Teaching students how (not) to lie, manipulate, and mislead, with information visualization. In S. A. Matei, N. Jullien, S. Goggins (Eds.), *Big Data factories: Collaborative approaches* (pp. 101–114). New York, NY: Springer.
- Mascarenhas, M. (2018). White space and dark matter: Prying open the black box of STS. *Science, Technology, & Human Values*, 43(2), 151–170. doi:10.1177/0162243918754672
- Massanari, A. (2018). Rethinking research ethics, power, and the risk of visibility in the era of the “alt-right” gaze. *Social Media + Society*, 4(2), 1–9. doi:10.1177/2056305118768302
- Metcalfe, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1). doi:10.1177/2053951716650211
- Nakamura, L., & Chow-White, P. (Eds.). (2012). *Race after the internet*. New York, NY: Routledge.
- Noble, S. U. (2018). *Algorithms of oppression*. New York: NYU Press.
- Oosterloo, S., & Van Schie, G. (2018). The politics and biases of the “crime anticipation system” of the Dutch police. In J. Bates, P. Clough, R. Jäschke, & J. Otterbacher (Eds.), *Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems* (pp. 30–41). Retrieved from http://ir.shef.ac.uk/bias/bias2018_proceedings.pdf#page=36
- Parks, L., & Starosielski, N. (Eds.). (2015). *Signal traffic: Critical studies of media infrastructures*. Urbana: University of Illinois Press.

- Qui, J. (2016). *Goodbye islave*. Chicago: University of Illinois Press.
- Scott, N., & Siltanen, J. (2017). Intersectionality and quantitative methods: Assessing regression from a feminist perspective. *International Journal of Social Research Methodology*, 20(4), 373–385. doi:10.1080/13645579.2016.1201328
- Todd, Z. (2015). Indigenizing the Anthropocene. In H. Davis & E. Turpin (Eds.), *Art in the Anthropocene: Encounters among aesthetics, politics, environment and epistemology* (pp. 241–254). London, UK: Open Humanities Press.

Further Reading

- D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. Cambridge, MA: MIT Press.
- Fotopoulou, A. (2020). *Feminist data studies: Big data, critique and social justice*. London, UK: Sage.
- Housley, W., Procter, R., Edwards, A., Burnap, P., Williams, M., Sloan, L., ... Greenhill, A. (2014). Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society*, 1(2), 1–15. doi:10.1177/2053951714545135

Mary Elizabeth Luka, PhD, is assistant professor of arts and media management in the Department of Arts, Culture & Media, UTSC, and cross-appointed to the Faculty of Information, University of Toronto, in Toronto, Canada. She is a critical communication and media scholar interested in creativity and emerging business models in cultural industries and urban development, and broadcasting, telecommunications, and culture policy. Recent publications include articles in the *Canadian Journal of Communication*, *Studies in Social Justice*, and *Social Media + Society*, and anthologies such as *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*.

Koen Leurs is assistant professor in gender and postcolonial studies at the Graduate Gender Program, Department of Media and Culture Studies at Utrecht University, the Netherlands. He is a critical internet researcher interested in migration, diaspora, youth culture, and urbanity. Leurs has recently published *Digital Passages: Migrant Youth 2.0. Diaspora, Gender & Youth Cultural Intersections* (Amsterdam University Press, 2015). He recently guest-edited special issues on “Connected Migrants” for *Popular Communication* and “Digital Connectivity and Forced Migration in(to) Europe” for *Social Media + Society*. He is the chair of the Diaspora, Migration & the Media division of the European Communication Research and Education Association (ECREA) and was the previous *Information, Communication & Society* annual special issue editor for the Association of Internet Research (AoIR). Currently, he is writing a monograph titled “Digital Migration.”