Absent Data: Engagements with Absence in a Twitter Collection Process

Katrine Meldgaard Kjær

Technologies in Practice, ETHOS Lab, IT University of Copenhagen
kakj@itu.dk

Mace Ojala

Technologies in Practice, ETHOS Lab, IT University of Copenhagen
maco@itu.dk

Line Henriksen

School of Arts and Communication, Medea Lab, Malmö University
line.henriksen@mau.se

Abstract

This paper considers the ways in which silences and absences are a central part of research that relies on automated data collection from social media or the internet. In recent years, automated data collection driven or supported research methods have gained popularity within the social sciences and humanities. With this increase in popularity, it becomes ever more pertinent to consider how to engage with digital data, and how both engagement and data are situated, messy, and contingent. Based on experiences with “missing” data, this paper mobilizes the framework of hauntology to make sense of what relationships may be built with missing data and how silences haunt research practices. Ultimately, we argue that it is possible to reimagine absent data not as a limitation but as an invitation to reflect on and establish new methods for working with automated data collections.

Kjær, Katrine Meldgaard, Mace Ojala, and Line Henriksen. “Absent Data: Engagements with Absence in a Twitter Collection Process.” *Catalyst: Feminism, Theory, Technoscience* 7 (2): 1–21.

<http://www.catalystjournal.org> | ISSN: 2380-3312

© Katrine Meldgaard Kjær, Mace Ojala, and Line Henriksen, 2021 | Licensed to the Catalyst Project under a Creative Commons Attribution Non-Commercial No Derivatives license

Keywords

Twitter, hauntology, absence, ghosts, data

Introduction: Partial Perspectives and Automated Data Collections

Figure 1. Email from institutional IT services notifying server owners of a power outage event.

This paper considers the ways in which absences and “missing” data are a central part of research relying on automated data collection from social media or the internet, and how this centrality might help us rethink and reimagine ideas of data and objectivity in this type of techno-scientific research. We argue that it is possible to reimagine apparently absent or missing data not as a limitation or failure but rather as a presence that co-constructs research processes, and which we therefore have an ethical obligation to respond to and engage with.

In recent years, automated data collection tools have gained popularity within research on and with social media, as so-called digital methods approaches, spearheaded by figures such as Richard Rogers (2013, 2019) and Noortje Marres (2017; Marres and Gerlitz 2016; Marres and Moats 2015), have highlighted the use of large datasets from social media platforms, collected via platform-specific application programming interfaces (API) in social research. However, critical data studies scholars such as danah boyd and Kate Crawford (2012) and José van Dijck (2014) have critiqued a “turn” towards big data methods in social research, and drawn attention to how media collection and analysis tools, as well as the datasets they create, are neither complete nor impartial. In related scholarship from science and technology studies (STS), Daniela Agostinho et al. suggest that

large data archives are fundamentally uncertain and marked by unknowability, error, and vulnerability (2019, 423); and Nanna Thylstrup et al. argue that “datafied knowledge production involves similar contingencies, limitations and complexities to previous forms of knowledge production” (2019, 1). This complexity makes well-known questions of epistemology, ontology, and method relevant in new ways, specifically within the digital realm. Indeed, as feminist philosophers of science and technology such as Donna Haraway (1988) and Karen Barad (2007) have long argued in relation to knowledge production, partial perspectives necessarily underpin all types of research; what is important is therefore not to seek to attain completion or neutrality as an implicit “gaze from nowhere” (Haraway 1988, 581), but rather to remain accountable for how and what one learns to see. With this, engaging the silences, absences, and what is *not* seen, also called agnotology (Proctor 1995; Schiebinger 2004), is an ongoing and important object of interest for feminist scholars when it comes to developing ways to think about how knowledge is created in collaboration with the digital. What is at stake here is therefore not the ability to fully include all perspectives, since this is impossible, but rather to develop an ethics of responsibility as one remains accountable for how one’s methods co-create the world they seek to understand. A sensitivity to the absent and the invisible is therefore an acknowledgement of how things could have been—and might be—different.

For research based on digital data collection from social media, engaging with absence entails examining what it means to take seriously that these research processes do not always yield the results they were intended to. Instead, they often involve silences and absences of data, whether due to glitches, breakdowns, and/or what the operators—including the busy, enthusiastic, and perhaps insufficiently (self?)trained social media researcher—might have missed. To examine these absences, our paper draws on empirical material and first-hand experience from a specific research project working with automated collection tools in order to map out Twitter debates on medicinal cannabis in Denmark. Taking a point of departure in the multiple breakdowns, outages, and disruptions that our collection was subject to during our research process, we argue that the silences brought on by techno-material possibilities and constraints of working with automated data collection tools in digital social media research are worthy of their own research spotlight. We investigate how the seemingly absent remain constitutive of any research results, in the sense that what cannot be measured—whether due to events such as glitches or simply the limitations of any research tool—shape how we know the world. The absent is therefore never fully absent, but a strangely paradoxical presence, and throughout this article we ask how one may conceptualize such present absences.

To unpack this question, we will mobilize the theoretical framework of Jacques Derrida’s “hauntology,” a portmanteau of *haunting* and *ontology* (Derrida 1994). We suggest that hauntology may offer a useful framework for conceptualizing

absent and silent data that nonetheless “haunt” the final data set and the research based on it. Hauntology forms part of a recent, more general scholarly interest in spectrality, sometimes referred to as “the spectral turn” (del Pilar Blanco and Peeren 2013). Not least, technological developments and increased digitalization have prompted scholars to turn to the ghostly and the spectral to theorize the complex, often hidden, and invisible workings of, for example, software (e.g., Chun 2011). While some scholars have applied Derrida’s work to the internet and questions of “data” (Blackman 2019; Fisher 2012, 2014), the extent to which hauntology can be useful in conceptualizing and theorizing what it means to do research in and on digitalized societies is still to be explored. This paper helps fill this gap and contributes to a body of scholarship within STS and critical data studies that is concerned with unpacking absences and silences in digital data. Our contribution to this discussion is novel because we argue that it does not make sense to approach these silences as indicators of “good” or “bad” data, nor to assume that the existence of silences are in themselves problematic. Rather, we argue that by using the lens of hauntology, it becomes possible to explore how these apparent absences are deeply relational and co-constructive, and therefore an inherent part of digital data research. Acknowledging traces of the absent within research results is the beginnings of what we, through Derrida, refer to as an ethics of living with ghosts rather than attempting to hide or ignore what cannot or/and should not be made fully present in order to give an impression of completeness.

Throughout the article, we share some of our internal email correspondences around a situation where we found our own research infrastructure to be haunted. By presenting email messages as empirical material, we attempt to account for our own partial perspectives, including our different roles and commitments to the project (see also Long et al. 2019). Our author group represents different scholarly fields, and we went into the project with a variety of roles, expectations, expertise, and entanglements: Katrine as a qualitative researcher and principal investigator of the medicinal cannabis project; Mace as part-time system administrator of locally operated Twitter Capture and Analytics Toolset (TCAT); and Line as researcher with analytical expertise with hauntology. Katrine and Mace began the data collection together, and Line was invited into the project in the analytical phase. Bradley et al. (2018) argue that much current research that seeks to merge humanities and technical research traditions for big data and digital data analysis end up with a division of roles that separates technical expertise and qualitative analytical work into discrete entities, thus limiting transformative potential for both parties. Similar challenges have been discussed within digital humanities (Terras et al. 2013). Academic interdisciplinarity, while often discussed and claimed, is difficult to perform; while much conceptual and philosophical work on the pros and cons of interdisciplinarity has been published, more concrete and specific accounts of the everyday and perhaps even mundane work with interdisciplinary collaboration and training are still largely missing

(Bornakke and Due 2018). By writing this paper together in a way that addresses different areas of expertise and expose traditionally hidden documents of research, we experiment with writing in ways that acknowledges our own partial perspectives visibly across lines of expertise and expectations.

We will first introduce how we have encountered silences in our datasets, primarily via a series of breakdowns and glitches in the data collection process of our research. Here we argue that these types of silences and absences are not unique to our data collection, but indeed characterize research based on automated data collection from social media more generally. We will then present Derrida's work on hauntology and argue that this framework can help us make sense of how to conceptualize and think *with* these absences, rather than dismissing them as failures or mistakes. Finally, we return to our absences and analyze how hauntology provides tools to rethink ideas of objectivity in research relying on automated data collection.

Initial Encounters with Absent Data

In January 2018, a four-year pilot program that permits doctors to prescribe medicinal cannabis started in Denmark. Medicinal cannabis is a controversial and contested subject globally (Zarhin et al. 2019), and given her research interest in the relationship between health controversies and media, Katrine asked Mace to assist in setting up a data collection to follow the debate about medicinal cannabis during the pilot program on Twitter. Collecting data from the social media platform was one part of a larger collection of data about medicinal cannabis debates from a range of sources, including newspapers, internet archives, and interviews. Although the ethics of relying on large companies for social media research data have been discussed (Cooky, Linabary, and Corple 2018), the relative ease of data access has nonetheless made Twitter a popular source for controversy mapping approaches. To follow the evolution of the debate, Katrine and Mace set up a collection process together targeting the hashtag #medicinskannabis plus a selection of related hashtags and keywords related to the controversy. The ongoing data collection discussed in this paper was set up on January 11, 2019 on an instance of the Twitter Capture and Analytics Toolset (TCAT) data collection software (Borra and Rieder 2014) hosted at our institution. In this paper, we draw on material from the first year and a half of this collection, which consists of approximately nine hundred tweets created by about five hundred unique users. When we began our data collection process, we were concerned with "capturing" data, aiming for a complete dataset from which to create an overview of the debate. Our ideas of "validity" of research were here tied to completeness and overview. However, what we faced was an urgent need in our research process to conceptualize our involvement with data that turned out to be apparently incomplete. We found that we needed a language and framework to think with the absences that characterized our datasets not as failures, but instead as a present actor that co-constructed our process. This

points to a research gap; while critical data scholarship has examined absences as being a part of data collections, there is still a need for work that informs how we might live with that absence not as mistake, but instead as a basic premise for working with data. We present a suggestion for one such framework by thinking with hauntings and ghosts as a part of digital research. Within this framework, the question of validity becomes situated, in that validity in datawork transforms into an imperative to account for absences in specific contexts.

Already within the first month of our data collection, we were confronted with unexpected absences when a first infrastructural breakdown occurred. This resulted in missing several weeks' worth of tweets:

Figure 2. An email of one of the authors asking another author for specimen of missing data event.

At this early stage our TCAT instance proved itself to be less than compliant, and we were confronted with the ways in which technology is an active participant in the development of what we can and cannot see in our data. Moreover, the specific breakdown was also a confrontation with how we do not always fully control or understand why collection software such as TCAT does not comply with our expectations. Figuring out why the software is not working is in itself an endeavor that requires imagination, professional vision (Goodwin 1994), and going beyond immediate assumptions. One may even argue that it is a labor of forensics and diagnosis. Indeed, beyond recognized and diagnosed cases of breakdown events, it is not always clear when an instance of TCAT is actually functioning as intended. For a relatively niche hashtag such as #medicinskannabis, it is possible that the hashtag simply is not used by anyone for days, meaning that there is no data to collect at that given time. When observing a single data collection within a TCAT instance, the distinction between days of non-activity of a technical infrastructure as opposed to days of non-activity in the topical debate may thus be unclear. In these ways, the research instruments and the social media platform are entangled with the phenomenon and co-constitute it. In the case of the issues with collecting hashtags and keywords relating to medicinal cannabis in early 2019, an iterative process of forensics revealed that the breakdown incident was part of a series of events that had already started before the medicinal cannabis research began: in the months

preceding the medicinal cannabis collection, disk space had run out on the server our TCAT was running on due to other research and teaching activities taking space on the same server infrastructure. This led to corruption of the database where the TCAT instance stores the tweets it collects. The medicinal cannabis research project had arrived in an already busy site of data infrastructure, with elements that were in varying states of maintenance.

The non-compliance our TCAT instance was exhibiting already in the early stages of the project, however, may well be considered an always-already inevitability of this type of research. Although Twitter is one of the most accessible social media platforms from which to digitally collect data, it is not clear which tweets are accessible via Twitter's tracking API (Borra and Rieder 2014; Cooky, Linabary, and Corple 2018). What *is* clear, however, is that the API does not provide access to all tweets made by users. For this access, Twitter requires financial compensation and control the process of collection themselves. We, like many other social media researchers, have chosen not to purchase data products from Twitter, as this raises new issues concerning the intersection of capital and social media research (see Cooky, Linabary, and Corple 2018). Moreover, drawing again on the central idea that digital collections are not neutral or complete, feminist and intersectional archival studies have argued that silences and absences about specific perspectives and voices often characterize digital archives, and that, as such, we need to work specifically with developing ways to conceptualize, engage, and think with them (see, e.g., Caswell and Mallick 2014; Cifor and Wood 2017). We cannot buy our way out of thinking with and about the significance and role of absences in digital data collections.

In our research project, insights regarding the technical and institutional conditions concerning the initial breakdown did not change the impact this breakdown had on the growing #medicinskannabis dataset. Later in the research process, the dataset came to be characterized by a second longer period of silence, occurring during the vacation season in the summer of 2019. Over the course of a year of data collection, as TCAT's own visualization of the dataset shows, almost three months were absent (Figure 3).

In this visualization of the dataset, absences and presences co-exist and co-shape each other: the non-present is not invisible, but rather a very salient part of the dataset and its reality. The visualization is akin to that of a heartbeat; the drops are as significant as the spikes and require analytical attentiveness on their own terms. The graph may point toward any number of causes, including the possibilities of the data collecting apparatus entangling with other systems on the shared infrastructure we were reliant on, or that there might have been pauses in the discussion on Twitter on the topic of interest. This made us ask questions about the ways in which absent data shape present data, including how we might imagine and theorize the impact of absent data on the finished dataset. After all,

if—as we and others suggest—glitches, absences, silences, and breakdowns are par for the course when working with technologies such as TCAT, we need analytical and conceptual frameworks that help us navigate these silences, invisibilities, and absences. A conceptualization of this sort requires a reorientation in thinking about ontology and what can legitimately be recognized as a part of a research project and its results. In order to gather tools to work with this reorientation, we turned to the imaginaries of hauntings and hauntologies, as well as their relationship with feminist theory on objectivity and partial perspectives in research processes.

Figure 3. A line chart showing daily Twitter data capture rates, with two notable gaps.

Figure 4. Email of one author reporting a login failure, which surfaces a server downtime event.

An Ontology of Haunting

Derrida coined the term *hauntology* in *Specters of Marx* (1994). According to María del Pilar Blanco and Esther Peeren (2013), *Specters of Marx* marked the early days of what some call the “spectral turn”—that is, an emerging field of multidisciplinary research that engages the figures of specters, ghosts, and hauntings to theorize the impacts of the invisible, the seemingly absent and the silent on the visible, the seemingly present and the audible. Research areas such as feminist theory, queer theory, and postcolonialism, among others, have a long tradition of theorizing and conceptualizing such silences and absences, not least as they relate to marginalized knowledges and lingering trauma of colonialism, for example. Spectralities have therefore been embraced by some theorists within these fields to conceptualize haunting traces, silences, and trauma (see, e.g., Castle 1995; Gordon 2008; McCormack 2014; Taylor 2019).

With hauntology, Derrida argues that traditional Western ontology primarily concerns itself with that which is present and immediate, suggesting that what can be said to exist is first of all that which can be seen, touched, or, in some way or another, measured and confirmed to be *here* and *now*. To exemplify how this affects traditional understandings of scientific objectivity, Derrida raises the ghost of the “traditional scholar,” a figure embodying the most rigid adherence to presence and immediacy as indicators of existence. Derrida writes,

A traditional scholar does not believe in ghosts—nor in all that could be called the virtual space of spectrality. There has never been a scholar who, as such, does not believe in the sharp distinction between the real and the unreal, the actual and the inactual, the living and the non-living, being and non-being...in the opposition between what is present and what is not, for example in the form of objectivity. Beyond this opposition, there is, for the scholar only the hypothesis of a school of thought, theatrical fiction, literature and speculation. (Derrida 1994, 12)

To the traditional scholar, objectivity concerns the establishing of what is present—and therefore relevant—and what is not—and therefore irrelevant. In this sense, Derrida’s traditional scholar’s understanding of objectivity can be related to what Donna Haraway calls the “God trick”—scientific objectivity as an imagined disembodied gaze from nowhere (1988, 584). Both the traditional scholar and the God-trick scientist understand scientific objectivity as the act of dispassionately observing a presence in order to establish where it belongs in the hierarchy of things—that is, as the act of functioning as a “modest witness” to the matters of facts of the world (Haraway 1997; Shapin 1984). Haraway argues that the modest witness is highly gendered, racialized, and classed; within the Enlightenment imaginary, which created the modest witness (for a more detailed engagement with the creation of modest witnessing, see Shapin 1984), only men of European

descent and from nobility were considered capable of extracting themselves from their emotions, bodies, and impulses, and thus of looking at the world as it was: static and unchanging, made immediate and present through the experiments within their labs. In other words, they embodied the ideal of the human—rational, conscious, objective—in ways that were considered largely impossible for their gendered, racialized, and classed others, who were deemed too close to nature, the body, and the animal to transcend the matter of the world and behold it from the disinterested position of the God eye. This creation of the objective scientist as white and male has been critiqued by, among others, decolonial and feminist scholars (see, e.g., Spivak 1988; Harding 2011; Wright 2015), including Haraway, who makes it clear that the God trick is just that: a trick, and an impossibility since any gaze, whether biological, technological, or both, must stem from somewhere and thereby be situational, a partial perspective. It is never distanced, it is never dispassionate, and the embodied position of the scientist co-creates the final results, meaning that much which is not straightforwardly visible, measurable, or even fully understood (thoughts, feelings, affects, memories, biases, etc.) form part of the final research results. They are, in a sense, haunted by present absences—by ghosts.

The traditional scholar, however, does not believe in ghosts, in the sense of that which is relational and therefore that which cannot be understood as either fully present or absent. He (*sic*) will only engage them in order to exorcise them, that is, explain them away or ultimately dismiss them as non-existent, as absent. According to Derrida, this dismissal of ghosts, of the present absent, complicates attempts at theorizing and addressing the impact of that which is not here and now (available to touch, measurement, and gaze) yet nonetheless strangely present. This might be the past, the future, or indeed the workings of tele- and information technologies.

Hauntology, to Derrida, is an attempt at theorizing and conceptualizing the impacts of the present absent, such as memories, speculation, techno-imageries. Although he did not live to fully theorize the internet through the lens of hauntology, other scholars interested in conversing with ghosts have since pointed out the ways in which digital media create and circulate specters (Blackman 2019; Fisher 2014). According to Wendy Hui Kyong Chun digital technologies have sparked imaginaries of omniscience and control through their promise of absolute transparency and the ability to turn the invisible visible—of communicating with and channeling the ghostly present absences controlling the machine (Chun 2011, 50). Indeed, Chun shows how tropes of magic and the supernatural saturate imaginaries of the workings of invisible systems whose effects are related to causes not always fully understood, and not entirely determined by their programming. As a medium, computers seemingly allow us to see the invisible and the hidden, while at the same time obfuscating that software does not magically open a window to pre-existing information, but

actively generates it (Chun 2011, 17). In alignment with practical programming knowledge, she argues that it is ultimately impossible to fully know the workings of software and that, instead of forcibly insisting on this (impossible) knowing, we must learn to take the lack of complete knowledge and transparency as “an enabling condition: a way for us to engage the surprises generated by a programmability that, try as it might, cannot entirely prepare us for the future” (Chun 2011, 54). We build on and extend this argument with reference to our specific empirical example where hauntings and hauntology may be put to critical use. By applying a hauntological lens, the question is not how to establish whether something is truly present or whether it is absent, but instead how to find traces of that which haunts presence, that which haunts the here and now. Through the figure of the specter and its hauntings, Derrida argues that the ethical imperative of the scholar is not to exorcise the ghost by asking that it either manifests itself or disappears, but to learn to live in the company of ghosts—that is, in the company of that which is neither fully manifest nor exorcised. With the trace of otherness and that which cannot be fully defined or known. This is living with the ghost as apparition; that which appears and disappears across vast distances, both temporally and spatially, without taking a final form or inhabiting a specific space. The apparition is a creature that troubles both the traditional scholar and Haraway’s scientist of the God’s-eye perspective in the sense that it cannot be fully seen and understood through a distanced, disinterested gaze, nor can it be kept at arm’s length. Further still, following Chun’s suggestion, the technologies applied to create the God’s-eye perspective do not offer transparency but generate invisibilities in order to make something else visible. They summon the ghosts they were meant to exorcise. Learning to live with this summoning, with these ghosts of undecidability and unknowability, is a challenge we wish to address through hauntology.

Living with Ghosts

Derrida’s argument that the “scholar of tomorrow” (Derrida 1994, 221) has an ethical obligation to learn how to converse with ghosts and stay in their company, rather than exorcise them, encourages us to confront and rethink established ideas that pertain to working with tools for automated data collection. The breakdowns made it clear that working with a tool such as TCAT involves a qualitative engagement first with data that may or may not have been missing, and second with the research instrument itself. This need for qualitative engagement showcases multiple aspects of the ways in which expectations towards digital tools and methods require constant reconfiguring. Most centrally, the breakdowns reveal that digital and automated data collections involve unstable relationships between what is known and not known. An investigation concerning how to live with these absences and the unknown therefore requires that we first understand that these are always already a part of the research process. In this section, we will return to the research process around collecting tweets on *#medicinskannabis* to analyze how absences came to be a presence

within this research process, and what the implications of this may be. To do so, we will return to our emails as data.

Figure 5. Email exchange about Twitter data collection downtime between the authors.

To Derrida, an “apparition” (1994, 157) is a ghostly visitation, a vision that appears and disappears suddenly, without the possibility of re-enacting the dis/appearance. Likewise, in our research process, we often did not know why, precisely, breakdowns occurred or what they affected, as they could be due to unpredictable, multiple, and overlapping institutional, technical, or personnel-related factors. While we engaged in a constant game of observation, we were left only with the haunting remains of an incomplete dataset. Already from the first breakdown, the absence of data formed part of our basic understanding of present data. This was very concretely represented in the gaps or dips in the graph of the collected number of tweets shown above (Figure 3), as the absences clearly confronted us here. However, the absences also became very present on account of their prominence in our discussions about the research project. Collectively, we had an ongoing orientation toward the absences or even the possibility of absences, as we checked in on the collection to ensure that it was still running and collecting as planned. The vast majority of our correspondences about the research project while collecting this data were about absences, either potential or actual. We also enrolled new actors into our research group by building software agents to keep sentry at the haunted sites where earlier apparitions had

visited. This ongoing orientation and attention toward absence entailed that absences had—both before and after they appeared in a dataset—become embedded within the research project itself. The breakdowns and breaks in data collection were invisible and silent until their effects were discovered, but they were nonetheless a constant presence on account of the collective anticipation of and reflections on them. The temporality of the breakdowns played a key role in our collective occupation with anticipating them: the suddenness with which absences in data collection appear and disappear, often leaving no clear indication as to why they came and went, required constant observation as well as attempted anticipation of when, where, and how breakdowns might happen. Yet, despite our efforts, there was a conundrum to breakdowns: they could only be located when it was “too late,” and then only as the haunting trace of an event long since passed.

Figure 6. One author reporting data collection to be down, another one responds by noting an institution-wide power outage event.

When we detected absences, they set in motion a new set of processes in the project, which aimed at returning the collection to previous operation. This post-apparition work of nudging, poking, and prodding became a central part of maintenance and continuation of the data collection process. The acts of nudging that were necessary to restart the data collection are acts of experimentation, as well as acts of material reparation and programming new software tools to patrol database infrastructure out of joint. However, even when the collection resumed, the marks left by the apparitions could not be erased or mended. These were now permanently in our datasets in the form of gaps or invisibilities, and the data that was “missed” during the ghostly visitations could not be re-collected; since neither TCAT nor Twitter data products at our disposal allow retrieval of tweets in the past, once data is identified as “missing,” it cannot become a part of the dataset. However, the missing data are in fact very visible in their not-thereness,

and they make their present absence known as they confront researchers with how they are not in control, not ultimately all-knowing, even despite attempts to be so.

The sudden activation of forensic and programming work, enrollment of software nightguards and the permanent gaps in the dataset all point to different ways in which absences or silences are not nothings. The absences do things: they set new labor in motion, involve new entities, and leave marks on the research. The absences are an active part of the research process and cannot simply be dismissed as mistakes, failures, solved problems, or limitations. Rather, absences, in all their ghostly presence, should be acknowledged as co-creators of the research process, with agency of their own. This acknowledgement also involves the recognition that control is not the prerogative of the human researchers on the project; collaborating with TCAT and automated data collection tools like it thus involves collaborating with the more-than-human.

Figure 7. Email of an author encouraging project documentation and announcing plans of designing monitoring instrumentation.

This role as active co-creator with agency and power of its own is clear from the ongoing and sometimes inevitably failed attempts to control and monitor the tool. Throughout the research process, we engaged in acts of reporting and trying to “tame” or make accessible the tool, to figure it out, and to prevent the ghostly visitations that nonetheless still occurred. Our orientation towards and our constant attempts to make TCAT monitorable speaks to the ways in which there is a constant lingering potential of disruption. This, in turn, challenges the idea that such tools are simply effective and neutral gateways to complete data collection; rather they are troublesome, active, and situated actors in the research process. However, as the above correspondence (Figure 7) shows, the ghostly visitations that occurred in our research process also set in motion new forms of collaboration, with the present absences as a central concern. In addition to the interdisciplinary collaboration the authors of this paper engaged in with the aim of making sense of the silent or absent in our data collection, they also inspired international and cross-disciplinary knowledge and code sharing concerning the

tool. What united these collaborations was an orientation toward precisely the absences, but also a care for the tool. We were not merely being troubled or annoyed by the absences in our monitoring; we checked in on them, we communicated about them, we worried about them. They became a site of intensity and affective flows, which created collaborative efforts both within and beyond our group.

Through these efforts we came to realize that ghosts have and will continue to visit, and breakdowns and bugs will occur. These ghosts force us to rethink ideas about how “good” science is produced and what it entails. Within the scientific tradition that Haraway and Derrida critique, science entails a specific kind of objectivity based on the ability to distinguish between what is present and what is not, what is actual and what is inactual, usually through the means of measurement and monitoring. However, Haraway (1988) argues that something will always escape this measurement and monitoring, sometimes because this “something” lies at the periphery of the gaze, sometimes because its existence eludes the ability of the eye to even notice it. During our ongoing attempts to monitor and observe the data, the ghostly visitations did indeed escape our gaze, but left a mark nonetheless. Rather than understanding this as a failure to achieve objectivity, we suggest that it indicates that this understanding of objectivity needs to be fundamentally reworked, also when it comes to work with automated digital data collections that can easily be presumed to be neutral or all-seeing. We therefore suggest that the figure of the haunting ghost and dis/appearing apparition may be applied or mobilized as means of acknowledging and attempting to stay with the silences, absences, and invisibilities in datasets without demanding that they be “solved,” dismissed, or asked to materialize as intelligible data.

Attending to these points thus involves a more general critique of the traditional understanding of “objectivity” in research. Feminist and intersectional research has critiqued the idea of objectivity as a detached view-from-above that results in knowledge generation unaffected by context for decades. However, such critiques remain controversial in other fields, not least due to the challenge they represent for the researcher’s position in relation to their research. These perspectives fundamentally destabilize traditional researcher positions as they reintroduce the researcher as embodied and situated, thereby arguing that the knowledge the researcher generates is equally embodied and situated. This idea departs radically from traditional positionings of the researcher as the all-knowing God-like traditional scholar observing from above, untainted and untouched by context—and the knowledge he (*sic*) produces as untainted and unaffected by context. Within this understanding of objectivity, research instruments are also necessarily neutral, acting only as vessels for the authority of the omnivoyant traditional scholar to gather and reproduce data and results. However, as Barad (2007) and others argue, instruments are also situated and affected, as are the

data they produce. An ethical reworking of ideas of objectivity in research, then, depends on an awareness of the situatedness of these instruments and the ways in which they create partial perspectives. Accounting for how they produce invisible, silent, or absent data and the circumstances of this production is one such way to situate the research. Emphasizing how the gaps or absences in data are co-constructive of the research process even in their invisibility is a central part of this rethinking objectivity.

In our research, the invisibilities in the dataset became very visible through our correspondences about them. Here, we brought absences into being by worrying about them, observing them, locating them, relating them to what was not missing, thinking about them, designing and programming for them. Rather than being not-there or not-noticed, the invisibilities in our dataset thus involved intensive labor, noticing, and ongoing negotiation. Throughout our research, we were thus collaborating with, fighting against, and working towards a sense of haunting; in the process, this haunting redirected our gaze and caused us to pay attention to the tools we were using in new ways; it confronted us with the liveliness of these tools, and it drew attention to basic methodological considerations concerning the nature of the absences we were, paradoxically, observing. The apparitions in our datasets could not be fixed—once this ghostly vision had visited our research, we could not undo it or cover it up by retrieving data to fill out the void. It had, instead, left a mark of very visible invisibility in our dataset, which materialized as a significant part of the dataset and of the research process itself. The void was a key site of collaboration and interdisciplinary dialogue as well as ongoing research design development. In this way, the invisible was a powerful present absence in the research process.

Conclusions

In this paper, we argue that it is possible to reimagine absent data not as a limitation but as a haunting presence that should not be ignored. While we started our project aiming to gain a complete overview of a given phenomenon via an “objective” dataset, we quickly found out that the dataset also came to include absences. This forced us to reflect on our own expectations towards data and prompted the need for conceptual and theoretical resources that would allow consideration of the ethics involved in relating to such absences. Our main object of concern, then, moved from thinking about how we could gain overview with the use of data, to how we could imagine living with the data we (did not) have. We thus moved from an orientation towards the instrumental use of present data to instead focus on the conceptualization of absent data. This article also reflects that movement: we argue, quite simply, that there is a need for more research on the ways in which absent data is present in research processes, and how we might think about and with this. In practice, confrontations with absence in our research process transformed the ways we approached questions of validity and data—

namely, towards an approach where, inspired by Haraway and Derrida, validity is tied to situatedness and learning to account also for what is *not* there.

In our research, absence activated various parts of our research group and our infrastructure and invited processes of reflexivity. The invisible and absent are not necessarily not-there and non-existent, nor do they invalidate the research. On the contrary, silences and absences are subtle reminders that the tools and technologies used for visualization do not merely measure a static world but are active parts of its ontology. They shape the world by establishing what can be measured and visualized in the first place (and therefore become worthy of note) and that which cannot (and therefore takes a backseat). Absences and silences in datasets thus point to the impossibility of the God's eye perspective of the traditional scholar—that is, the impossibility of the researcher as a neutral all-knowing eye. Absent data also draw attention to the ways in which digital social media research is always-already partial and shaped by the situated perspectives of the materiality of research tools and practices—a point that challenges common assumptions of big and digital data-based social media research as producing coherent, complete, objective, and/or neutral research datasets (see also Cooky, Linabary, and Corple 2018; boyd and Crawford 2012; Chun 2009). It is clear that research on and with social media that relies on automated data collection tools needs to grapple with invisibilities and what appears to be non-data—or at least non-visualizable data—not least as these silences are becoming better documented within this type of research.

With this paper, we suggest that it is crucial to find ways to conceptualize and theorize that which escapes visualization and thus also the God's eye of the traditional scholar. We suggest that mobilizing hauntology and its figure of the specter may provide a way to acknowledge the presence of absences in datasets without demanding that they materialize—that is, take a steady form as “recognizable” data in a given dataset. This points towards our initial question about living with absence in digital research and how researchers may concern themselves with and perhaps even embrace silences in datasets by rethinking how they structure different modes of relationality between researchers and that which is researched.

This relationality is the ethical core of hauntology: to live with ghosts, in their company, without demanding that they take a steady, knowable form that can then be measured, explained, and understood. For a researcher, to learn to live with ghosts thus entails a sensitivity to the ways in which data will always be haunted by a present absence. Such hauntings are reminders of the otherness at stake in the material under scrutiny: that there is no God's eye perspective, only multiple, partial, contradictory perspectives, some of which belong to the ghosts themselves, the voices that were not “captured.” Informed by our own

experience, we encourage researchers to communicate with the haunting ghosts in their research, rather than attempt to exorcise them.

Acknowledgments

Thank you to the editors and anonymous reviewers for their thoughtful suggestions. Research for this article was supported by a grant from the Carlsberg Foundation CF18-0863 "Controversial Healing. Making sense of medicinal cannabis debates."

References

- Agostinho, Daniela, Catherine D'Ignazio, Annie Ring, Nanna Bonde Thylstrup, and Kristin Veel. 2019. "Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive." *Surveillance & Society* 17 (3–4): 422–41. <https://doi.org/10.24908/ss.v17i3/4.12330>.
- Barad, Karen. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.
- Blackman, Lisa. 2019. *Haunted Data: Affect, Transmedia, Weird Science*. London: Bloomsbury Academic.
- Bornakke, Tobias, and Brian L. Due. 2018. "Big–Thick Blending: A Method for Mixing Analytical Insights from Big and Thick Data Sources." *Big Data & Society* 5 (1). <https://doi.org/10.1177/2053951718765026>.
- Borra, Erik, and Bernhard Rieder. 2014. "Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets." *Aslib Journal of Information Management* 66 (3): 262–78. <https://doi.org/10.1108/AJIM-09-2013-0094>.
- boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5): 662–79. <https://doi.org/10.1080/1369118X.2012.678878>.
- Bradley, Adam James, Mennatallah El-Assady, Katherine Coles, Eric Alexander, Min Chen, Christopher Collins, Stefan Jänicke, and Dadiv Joseph Wrisley. 2018. "Visualization and the Digital Humanities." *IEEE Computer Graphics and Applications* 38 (6): 26–38. <https://doi.org/10.1109/MCG.2018.2878900>.
- Castle, Terry. 1995. *The Apparitional Lesbian*. Rev. ed. New York: Columbia University Press.
- Caswell, Michelle, and Samip Mallick. 2014. "Collecting the Easily Missed Stories: Digital Participatory Microhistory and the South Asian American Digital Archive." *Archives and Manuscripts* 42 (1): 73–86. <https://doi.org/10.1080/01576895.2014.880931>.
- Chun, Wendy Hui Kyong. 2009. "Introduction: Race and/as Technology; or, How to Do Things to Race." *Camera Obscura: Feminism, Culture, and Media Studies* 24 (1 (70)): 7–35. <https://doi.org/10.1215/02705346-2008-013>.

———. 2011. *Programmed Visions: Software and Memory*. Cambridge, MA: MIT Press.

Cifor, Marika, and Stacy Wood. 2007. "Critical Feminism in the Archives | Journal of Critical Library and Information Studies." *Critical Archival Studies* 1 (2).

<https://doi.org/10.24242/jclis.v1i2.27>.

Cooky, Cheryl, Jasmine R. Linabary, and Danielle J. Corple. 2018. "Navigating Big Data Dilemmas: Feminist Holistic Reflexivity in Social Media Research." *Big Data & Society* 5 (2). <https://doi.org/10.1177/2053951718807731>.

Derrida, Jacques. 1994. *Specters of Marx: The State of the Debt, the Work of Mourning and the New International*. New York: Routledge.

Fisher, Mark. 2012. "What Is Hauntology?" *Film Quarterly* 66 (1): 16–24.

<https://doi.org/10.1525/fq.2012.66.1.16>.

———. 2014. *Ghosts of My Life: Writings on Depression, Hauntology and Lost Futures*. Zero Books.

Goodwin, Charles. 1994. "Professional Vision." *American Anthropologist* 96 (3): 606–33. <https://doi.org/10.1525/aa.1994.96.3.02a00100>.

Gordon, Avery. 2008. *Ghostly Matters: Haunting and the Sociological Imagination*. Minneapolis: University of Minnesota Press.

Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99.

<https://doi.org/10.2307/3178066>.

———. 1997. *Modest_Witness@Second_Millennium.FemaleMan_Meets_OncoMouse*. New York: Routledge.

Harding, Sandra, ed. 2011. *The Postcolonial Science and Technology Studies Reader*. Durham, NC: Duke University Press.

Long, Ziyu, Jasmine R. Linabary, Patrice M. Buzzanell, Ashton Mouton, and Ranjani L. Rao. 2019. "Enacting Everyday Feminist Collaborations: Reflexive Becoming, Proactive Improvisation and Co-Learning Partnerships." *Gender, Work & Organization* 27 (4): 487–506. <https://doi.org/10.1111/gwao.12421>.

Marres, Noortje. 2017. *Digital Sociology: The Reinvention of Social Research*. Cambridge: Polity Press.

Marres, Noortje, and Carolin Gerlitz. 2016. "Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology." *The Sociological Review* 64 (1): 21–46. <https://doi.org/10.1111/1467-954X.12314>.

Marres, Noortje, and David Moats. 2015. "Mapping Controversies with Social Media: The Case for Symmetry." *Social Media + Society* 1 (2).

<https://doi.org/10.1177/2056305115604176>.

McCormack, Donna. 2014. *Queer Postcolonial Narratives and the Ethics of Witnessing*. London: Bloomsbury Academic.

Pilar Blanco, María del, and Esther Peeren. 2013. "The Spectral Turn." In *The Spectralities Reader: Ghosts and Haunting in Contemporary Cultural Theory*, edited by María del Pilar Blanco and Esther Peeren, 31–36. Bloomsbury Academic. E-book.

Proctor, Robert N. 1995. *Cancer Wars: How Politics Shapes What We Know and Don't Know about Cancer*. New York: Basic Books.

Rogers, Richard. 2013. *Digital Methods*. Cambridge, MA: MIT Press.

———. 2019. *Doing Digital Methods*. Cambridge, MA: MIT Press.

Schiebinger, Londa. 2004. "Feminist History of Colonial Science." *Hypatia* 19 (1): 233–54. <https://doi.org/10.1111/j.1527-2001.2004.tb01276.x>.

Shapin, Steven. 1984. "Pump and Circumstance: Robert Boyle's Literary Technology." *Social Studies of Science* 14 (4). <https://doi.org/10.1177/030631284014004001>.

Spivak, Gayatri Chakravorty. 1988. "Can the Subaltern Speak?" In *Marxism and the Interpretation of Culture*, edited by Cary Nelson and Lawrence Grossberg, 271–313. Basingstoke: Macmillan Education.

Taylor, Leila. 2019. *Darkly: Black History and America's Gothic Soul*. Illustrated Edition. London, United Kingdom: Repeater.

Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte, eds. 2013. *Defining Digital Humanities: A Reader*. New York: Routledge.

Thylstrup, Nanna Bonde, Mikkel Flyverbom, and Rasmus Helles. 2019. "Datafied Knowledge Production: Introduction to the Special Theme." *Big Data & Society* 6 (2). <https://doi.org/10.1177/2053951719875985>.

Van Dijck, José. 2014. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12 (2): 197–208. <https://doi.org/10.24908/ss.v12i2.4776>.

Wright, Michelle M. 2015. *Physics of Blackness: Beyond the Middle Passage Epistemology*. Minneapolis: University of Minnesota Press.

Zarhin, Dana, Maya Negev, Simon Vulfsons, and Sharon R. Sznitman. 2019. "'Medical Cannabis' as a Contested Medicine: Fighting over Epistemology and Morality." *Science, Technology, & Human Values* 45 (3). <https://doi.org/10.1177/0162243919862866>.

Author Bios

Katrine Meldgaard Kjær is an assistant professor at the IT University of Copenhagen. Her research focuses on critical data studies, digital methods and health discourses.

Mace Ojala is science and technology studies scholar focusing on cultures of computer software. Mace works as a teaching assistant and lecturer at IT University of Copenhagen.

Line Henriksen is a postdoctoral researcher at the School of Arts and Communication, Malmö University. She publishes on the topics of hauntology, monster theory, digital horror stories, and experimental methods.