Mobilizing Social Media Data: Reflections of a Researcher Mediating between Data and Organization

Adriana Alvarado Garcia
adriana.ag@ibm.com
IBM Thomas J. Watson Research Center
Yorktown Heights, NY, United States

Marisol Wong-Villacres
lvillacr@espol.edu.ec
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador

Milagros Miceli
m.miceli@tu-berlin.de
Technische Universität Berlin, Weizenbaum Institute, DAIR Institute
Berlin, Germany

Benjamín Hernández*
Oak Ridge National Laboratory
Oak Ridge, TN, United States

Christopher A. Le Dantec
ledantec@gatech.edu
School of Interactive Computing, Georgia Institute of Technology
Atlanta, United States

ABSTRACT

This paper examines the practices involved in mobilizing social media data from their site of production to the institutional context of non-profit organizations. We report on nine months of fieldwork with a transnational and intergovernmental organization using social media data to understand the role of grassroots initiatives in Mexico, in the unique context of the COVID-19 pandemic. We show how different stakeholders negotiate the definition of problems to be addressed with social media data, the collective creation of ground-truth, and the limitations involved in the process of extracting value from data. The meanings of social media data are not defined in advance; instead, they are contingent on the practices and needs of the organization that seeks to extract insights from the analysis. We conclude with a list of reflections and questions for researchers who mediate in the mobilization of social media data into non-profit organizations to inform humanitarian action.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; *Empirical studies in HCI*; *Ethnographic studies.*

KEYWORDS

social media data, user-generated content, NGO, non-profit organizations, data work, mobilizing data, organizational practices, data experts, humanitarian context

ACM Reference Format:
Adriana Alvarado Garcia, Marisol Wong-Villacres, Milagros Miceli, Benjamín Hernández, and Christopher A. Le Dantec. 2023. Mobilizing Social Media Data: Reflections of a Researcher Mediating between Data and Organization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3580916

*Also with NVIDIA.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580916

1 INTRODUCTION

Big data analysis is often presented as a desirable and promising pathway for non-profit organizations in the humanitarian context to determine when to act, what support to provide, and how to report their success back to their funders [26, 29, 34, 98]. However, such organizations face many problems to datify their work: their data collection process is often determined by outside demands (e.g., funders) [29, 117], they struggle to secure the human and technological resources needed for developing and sustaining a culture of data [32, 37, 84], and important levels of support to develop strategic thinking that can lead them to turn data into actions [37, 84].

Research in disciplines closely related to the field of Human-Computer Interaction (HCI) such as crisis informatics and computational social science, suggests that the use of social media data can help organizations navigate these issues. This type of data can facilitate abundant data collection [78, 80], seems to provide access to near real-time voices that can allow for faster deployment of support [54, 71], and contains nuanced evidence that can help non-profit organizations to hold institutions and governments accountable [18, 19, 116].

However, scholars within Critical Data Studies have issued warnings about the use of social media data for devising actions that affect human networks and communities [25, 45, 86, 104, 122]. Besides the bias that all data analytical processes via machine learning (ML) models introduce [41, 57, 92, 105, 112], social media data cannot guarantee that the perspectives and needs it highlights are representative of those who are in need of support [45, 68, 74, 96, 104, 122]. Further, it is unclear how organizations in the humanitarian context can derive actionable value from these data. To be used, social media data is taken out of its site of production; thus, it is stripped out of its meaning and its potential to generate comprehensive knowledge of the situation on the ground [36, 45]. Despite these problems, funders, researchers, and other decision-making actors continue to champion computational techniques that can mainstream the use of social media data for informing humanitarian action [17–19, 58, 65, 66, 131, 133]. It becomes, thus, critical to better understand *how non-profit organizations doing humanitarian work engage with the process of analyzing social media data to make effective, valuable decisions on the ground*. Such an understanding requires in-depth explorations of the challenges that these organizations face, how they handle them, and how these negotiations impact their initial data-related goals.

In this paper, we respond to this need and present an investigation into how an initiative within a large, transnational non-profit organization mobilized social media data to determine the capacities that citizens and governmental institutions in the city of Mexico were resorting to during the COVID-19 health crisis. *Mobilizing* social media data means moving such user-generated content from its place of origin—the social media platforms—into a different context. Human interpretation and situated data practices are at the core of this process [36, 101, 106]. Through ethnographic fieldwork and self-reflexive accounts, we provide a view into a project within the organization's process from the perspective of the first author who, thanks to an internship with the organization, worked as a data-expert researcher, mediating the relationship between the organization and its data-based needs and goals. As Erete et al. point out, the relationship between non-profit organizations and data experts is "critical to the use, consumption, and interpretation" of data [56].

Our findings describe different moments in the relationship between the organization and the data expert in her role as mediator between an organization and the data. Those moments involve practices and decisions related to *(1)* the definition of the dataset, *(2)* the creation and establishment of the ground truth data, *(3)* the sensemaking of the data, and *(4)* the extraction of insights and value. In showing how the decisions involved at each stage shaped the data and the insights obtained from the analysis, our findings extend previous research on the implications of datification for organizations [13, 28, 56, 127] to the context of social media data.

Drawing on our analysis, we discuss three critical assumptions related to the use of social media data in the context of non-profit organizations and propose a set of questions for data experts to consider when facing these assumptions. *First,* we contest that the use of social media data is straightforward, fast, and less expensive than using other types of data. As our findings highlight, the careful and reflexive mobilization of social media data requires a rather considerable amount of resources that not every organization has at its disposal. *Second,* we demonstrate that the assumed objective nature of findings developed from social media data is not necessarily the main value those data can provide to organizations. Our analysis shows that there is much richer value in the negotiations between non-profit staff and data experts, for it is there when important questions and reflections about the problem context emerge. The role of the data expert as a mediator between an organization and the data is, thus, crucial for helping the organization recognize and appreciate these unexpected forms of value. *Finally,* our analysis challenges the belief that the effectiveness of the data expert's work is based on the computational methods they use. It is, in fact, the researcher's commitment to deliver contextualized insights what allows for organizations to reflect on the decisions they can make. The data expert, thus, needs to develop a critical view over their own disciplinary practices and continuously reflect on how their positionality affects their analytical abilities and decisions.

The analysis and discussion we present in this paper offer a two-fold contribution to the line of research that explores the role of Information and Communication Technologies (ICTs) and data work in organizations carrying humanitarian interventions [12, 13, 26, 28, 34, 80, 83, 98, 119] — especially work that has identified specific barriers that humanitarian organizations face to integrate data into their processes and seek to offer these organizations feasible and reliable alternatives. First, in showing the decisions involved at each stage, we expand the existing understanding of the data expert's role [16, 75, 100–102] and the organizations' datification process [7, 23, 118, 126] to include specific considerations on the use of social media data which is often not generated with humanitarian goals in mind. Second, our discussion of assumptions (Section 6) contributes to the growing body of research on recommendations and frameworks for data experts to work towards social good focusing on data work [35, 43, 53, 67, 97]. In particular, the questions we propose in the discussion section can help data experts in their self-reflection journey as they delve into data mobilization processes, contributing thereby to discussions on how humanitarian organizations can navigate the pressure to innovate via datification [26, 80].

2 RELATED WORK

2.1 Use of Data within the Context of Non-Profit Organizations

The pressure for non-profit and humanitarian organizations to know where and how to invest their resources is high; ensuring that their actions can effectively attend to communities' most acute needs is critical to support society's well-being [56, 98, 124] and to organizations' own subsistence. Most of these organizations' operations depend on philanthropic and governmental funders who demand evidence that their funds are put to optimal use [26, 29, 56, 98, 124]. Motivated by the success of big data use in the private sector, various important organizations and actors in the non-profit and humanitarian field—including the World Humanitarian Summit [26, 80], external funders seeking accountability [38], and digital humanitarians [34]—are increasingly positioning the tracking, collection, and analysis of large datasets as critical operations for non-profits to satisfy their decision-making needs [26, 29, 38, 80, 98].

Research has shown that data-driven decision-making can improve the "performance, output, and productivity" [29] of organizations in the for-profit and public sectors [26, 29, 38, 56]; contexts in which big data are considered "the foundation for ongoing productivity, innovation, and competitiveness" [26]. Building on these success stories, organizations in the nonprofit and humanitarian sector see in big data analytics a highly promising path towards optimizing their operations and results [38, 56, 70, 109], leading them to "more productive and empowered decisions" [28] and enabling them to "respond to funders' performance measurement mandates" [38].

For humanitarian, non-profit organizations, deriving value from data, however, is far from seamless. For starters, identifying the data needed for engendering insights can be a struggle; in addition to having to consider their donors' data collection demands, non-profits need to choose from a vast as well as subjective, biased, and flawed amount of data [98]. Abstracting actionable knowledge from data demands organizations to recognize that data and instrumentation alone cannot fully reflect conditions "on the ground" [26, 33, 38]. In their exploration of how organizations in a coordinated immigration relief effort would perceive an automatic analysis of qualitative organization-generated data [81], for example, the authors found that organizations expected data to engender "compelling examples reflecting the alleviation of suffering through the assistance programs." However, members of the organization were highly disappointed when learning that the analysis only provided a description of assistance operations. To be able to hold realistic expectations about data and complement data lessons with other sources, organizations need to invest in developing a culture of data that assists organization staff in reaching a consensus about how data can be valuable to them [26, 84].

However, securing the technological resources and skilled staff [27, 29, 38, 56, 83], and even autonomy of decision-making [27, 29, 38] that such a culture needs can be quite challenging for these organizations[27, 29, 56, 72, 84, 124, 125]. Technology resources and skills, for example, are either scarce for many of these organizations [27, 29, 38, 56] or existent but underutilized due to inappropriate technology and/or policy design [38]. While many organizations navigate these issues via artful informal practices (e.g., assembling of patchworks of information systems [125]), their ability to plan and implement the technical side of data collection and management is still often impaired [29]. Further, even if organizations are eventually capable of overcoming technological barriers, many lack the autonomy and strategic thinking for using data beyond meeting short-term funding requirements [29]. These barriers feed a cycle of disempowerment, flooding organizations with data and preventing them from engaging in data practices that can lead them to effective decision-making [29, 119].

In this paper, we lay our focus on user-generated data, which is often presented as an alternative for organizations to navigate datification challenges. These data can give an easy access to massive amounts of evidence about people's expressed needs [34, 80, 88]. While promising, the nature of these data also hints at an important risk of further amplifying the issues that organizations already face. User-generated data is highly disconnected from its site of collection, and thus, it is prone to provide not only limited but highly misguided insights. As Maitland et al., found, rich qualitative data—either user-generated or organization-collected—entered to an analysis process from points far away from their context of origin risks "losing valuable contextual insights and silencing local experts" [81]. Emerging research has already began to identify the extend of user-generated data limitations when trying to provide lessons for humanitarian contexts. Burns et al.found that crowsourced data for crisis relief efforts' portrayal of local communities and their needs did not align with how these communities saw themselves [33]. Alvarado-Garcia et al. identified that, while organizations and activists can find valuable knowledge about citizens' perspectives during a humanitarian crisis on social media pages, they struggle to turn this data into verifiable evidence that can augment official data about the crises [12]. As they conclude in a later work, systematically turning social media data into evidence for guiding humanitarian actions requires "a robust method of collecting, analyzing, and cataloging social media data" [11].

Aware of these limitations, scholars from disciplines such as Science and Technology Studies (STS) and Critical Data Studies have called into question the belief that, given its abundance, user-generated data is complete [45, 122]. As [26, 33] explain, the process of searching meaning and patterns in this type of data, obfuscates "the institutional and community-based processes and limitations that frame the types of data produced and the representational strategies espoused." Thus, these scholars urge organizations engaging in data analysis—especially those seeking evidence to inform community and policy interventions—to reconsider how their decision-making process for determining what is of value about data and why, happens [26, 33]. This entails, engaging in more in-depth explorations about organizations' experiences as they decide what data to collect vs others, how they can recover and document the data's context of origin, how they unearth relations form data, how they negotiate data, and how they make sense of this understanding for taking action [26, 29, 33, 70, 73]. Given the critical role that the human intervention of skilled staff has in the construction and interpretation of data [25, 45, 86, 101, 104, 122], a path moving forward is to engage in the needed organizational explorations from the perspective of the data expert [56, 73]. Amongst important questions to explore when unpacking such a perspective are the factors that shape the experience of these individuals [73], the collaborative practices they engage in when working with the rest of the organization [56], and how data-driven demands shape these work practices and organizational identity as a whole Bopp et al. [29].

In this paper, our main concern is exploring the experience of researchers that adopt the role of a data expert, mediating how social media data becomes legible for non-profits. In particular, we are interested in understanding how the organizational practices and needs of stakeholders, who might benefit from the insights of social media data, and the mediators' human interpretation influence the selection and interpretation of that data. Therefore, our investigation is framed by the following research questions: *(RQ) What are the implications of mediating how NGOs mobilize social media data from their site of production to inform institutional decisions for addressing humanitarian crises?, (RQa) What are the practices that mediators engage in when supporting NGOs?*

2.2 Social Media Data and Human Interpretation

Studies on datification and institutions agree that the full promise of promoting data collection and analysis lies in the data-related process ability to produce value for the institution and its different stakeholders [26]. However, such a promise is underpinned by assumptions that need careful unpacking. Sharma et al., for example, explain that institutions and even researchers tend to hold a strong belief that "if the quality of decisions can be improved by analytics, then deriving value from those decisions is trivial" [115]. Other common assumptions are that the heterogeneity and massiveness of data ensures contextual relevance and sufficient granularity for deriving value [26] and that "instrumentation alone is sufficient to deliver value" [77]. As STS scholars have increasingly argued, these assumptions are dangerously simplistic: analytic systems and algorithms that aim to shape user experience and economic and labor decisions depend on human interpretation and socio-cultural values [51, 92, 106, 108], and thus, the inferences they produce by no means can be complete or objective [4, 31].

The human role in deriving value from data analytics begins even before the data reach the computer [45, 110] and is constantly present across all work practices in data science, including, discovery, capture, curation, design, and creation [16, 101]. The definition of the "dependent variable" used to train and validate ML models—the so-called ground truth data—for example, demands human ascription of the meaning—and thus, the value—behind the data through labelling and annotation [102]. These processes are typically based on the assumption that for each data point and annotation instance there is a single right answer [20]. This way, ground-truth labels synthesize the often infinitely more complex realities comprised in each data point to make them "readable" in computational terms [55]. Further, it determines what others should see as the objective truth, filtering out everything that is not consider of value; as argued by [53], "once a [classification] system is in place, it becomes naturalized as 'the way things are.'". However, ground truth definition and construction is far from being an objective process; it is a human endeavor [21, 102], one that entails negotiation [107], and is constrained by organizational practices, business goals, and existing technologies [93, 94, 99, 132]. Further, situated and hierarchical structures and power asymmetries heavily shape the process. For example, for the data annotators, what guides their interpretation of data is "the epistemic authority of managers and clients" rather than what is contained in the data [91, 93].

A growing body of STS and HCI research has demonstrated that decision-making practices determining the value of data have important ethical and social ramifications such as injustices in algorithmic systems, extractive data practices, and risks for impacting the resources of vulnerable populations [97]. Responding to the critical need for further understanding the data workers' practices that lead to these issues, existing research has highlighted that data science workers' decisions often respond to the many tensions they face across the analytics process. For example, they often rationalize and decompose data so as to reduce tensions when defining and enabling the data workflow [101, 107]. Further, they emphasize value on aspects such a data quality, scale and scope of data, and data uniqueness to maximize benefits for various stakeholders [64]. Those involved in the datification pipeline also struggle to align their views on how to value data or derive value from it. In the health sector, for example, Thakkar et al. [120] identifies three points of common misalignment: how to improve data quality, how to contextualize data, and how to turn data insights into actions.

Developing recommendations and methodologies for assisting data practitioners in their decision-making process as they develop what Passi and Jackson [106] define as a "data vision", or "the ability to organize and manipulate the world with data and algorithms", becomes thus critical. To that end, recent work has proposed a wide range of actions [35, 43, 67, 97, 114]. Scheuerman et al. [114], for example argue for a closer look at the trade-offs that highlighting one value over others can have when dealing with data (e.g., efficiency at the expense of care). Ismail and Kumar [67] recommend for practitioners to engage in more transparency-related practices towards their values and goals and especially expose those that are common for helping produce more balanced ML datasets. A growing group of academics draw from a decolonial critique approach to propose different pathways for reforming epistemic practice [43, 97]. Couldry and Mejias [43] propose to decolonize data relations via six social-good oriented tasks: reframing what data is for, restoring well-being, naming alternative world views, gendering, protecting, and creating new forms of social relations. Mohamed et al. [97] warn practitioners against seeking a conclusive method for analyzing the ethical and social harms of a particular technology. Instead the authors propose three tactics for furthering research and action in regards to AI and datification: working towards a critical practice, establishing reciprocal engagements and reverse pedagogies with those in the periphery, and renewing relations with political communities.

An area of data use that requires more of such a critical examination is the role of human interpretation and subjectivity in how institutions such as NGOs can derive value from social media data. The large amount of social media data being produced by the second for various different topics has shown important potential to inform decision-making in various fields including mental health [22, 49, 113, 121], public safety [59, 82, 130] and humanitarian response [8, 58, 131]. Existing work on how to derive valuable insights from social media data, however, has been limited to the evaluation of computational techniques, particularly machine learning, for inferring new meanings about people's behavior across contexts [22, 39, 48, 50, 113, 128, 129]. Mobilizing user-generated content, such as social media data, to an institutional context and abstracting then value from it, entails removing the data from its platform and context of origin into a different one [36, 45]. Such a shift from sites of practice demands a purposeful transformation of the content legibility which entails a series of interpretation-informed decisions, including which attributes of social media to consider, which to ignore, how attributes associate with each other and with human behavior [11, 104, 122]. Outside of the scope of social media data research, previous research on social science and humanities has examined the obstacles and implications of data sharing and re-use across scientific disciplines and how these challenges are bound to the context of data production [36, 123]. Further, this research has proposed methodologies and frameworks such as *data journey* to examine how data travel from sites of production into sites of processing and later into markets of use and re-use across diverse disciplines and institutions [24, 76].

Drawing from this previous work, in this paper we explore the particularities of social media data re-use in the context of an *initiative* within a non-profit organization assisted by a mediating data expert, seeking to provide effective, valuable support to communities in need. Therefore, our third research question is: (RQb)*How do mediators and NGO teams negotiate what social media data counts as evidence for their work and what challenges do they encounter?* Specifically, we explore how the culture of objectivity of the *initiative* as well as the mediating data expert's commitment to approach data from a community perspective, shaped the process of *defining, negotiating, and interpreting* data, including ground-truth data, to inform future interventions. With this reflection, we aim to provide a more nuanced and detailed understanding of the subjectivities embedded in the process of mobilizing social media data from their place of origin to institutional contexts.

3 CONTEXT AND RESEARCH SITE

In December 2019, the World Health Organization announced the existence of the infectious disease COVID-19, caused by the SARS-CoV-2 virus, after an outbreak occurred in the Chinese city of Wuhan [1]. In Mexico, the first confirmed case happened in Mexico City during the last week of February, and the first death from this disease in the country occurred on March 18, 2020 [6, 52]. The Government of Mexico, in coordination with the Ministry of Health, implemented various measures to prevent and control infections, including a monitoring system for the regulation of the use of public space according to the risk of contagion of COVID-19 [2]. At the time of initiating this research, on February 2020, Mexico City was on its highest alert level – only essential economic activities were allowed, which meant the suspension of cultural events and the closure of schools and retail outlets [44].

It is within this context that the first author collaborated for nine months in a project with the Accelerator Lab Mexico or AccLab-Mx initiative,[1] which belongs to the Accelerator Lab Network of the United Nations Development Programme (UNDP), the United Nations' lead agency on international development [3]. The Accelerator Lab Network consists of 91 labs that support 115 countries and focuses on working closely with local stakeholders to identify community-level solutions and new ways of working that have the potential to accelerate learning on what works and what does not for new sustainable development approaches. In an internship with the AccLab-Mx held between January 2020 to March 2022, the first author participated in field research and facilitated social data analysis processes.

In May of 2020, the AccLab-Mx launched the project *COVID-19 Social Inventory* [2] as part of their overall pandemic response. The AccLab-Mx's objective was to learn from local, citizen-based initiatives and identify insights that might inform the relationship between these initiatives and existing social capital [90]. In the following subsections we describe the scope of *the project* as well as our role within it. The following descriptions are not a report of the methods used in this paper. Rather, they explain how our investigation situates in the specific context of the the AccLab-Mx and *the project* it conducted. A description of the method that we used in the "meta" analysis that we report in this paper follows in Section 4.

3.1 The Project

The COVID-19 Social Inventory project consisted of identifying the existing capacities across communities in Mexico City that could support responses to the COVID-19 health crisis [90]. To that end, *the project* focused on identifying and analyzing citizen-led—individual or collective—initiatives already responding to the crisis. Specifically, the AccLab-Mx focused on two types of analysis of these initiatives. Following [9, 10], it sought to unpack the social capital behind citizen-led initiatives; three different types of social capital were the focus of this analysis: "bonding," which connects people within a community; "bridging," which enables connections across communities; and "linking," which connects communities with government. The AccLab-Mx also wanted to analyze how these initiatives reflected the different capacities that each *alcaldía* in Mexico City was using for responding to the crisis.[3]

The first step towards analyzing initiatives was to identify them. The AccLab-Mx created an online survey, inquiring about the purpose, the organizer, and the scale and scope of the initiative. The survey was distributed during three weeks with the help of 37 collaborators from a partner organization and helped to identify 172 citizen-led initiatives. From then on, survey responses consistently declined. Since the survey was being distributed through the personal network of the volunteers, the AccLab-Mx concluded that they had already exhausted all human resources and decided to explore other forms of data collection.

After internal discussions between the AccLab-Mx, its partner organization, and the first author about possible paths moving forward, the AccLab-Mx decided to explore Twitter data as an option. The first author was in charge of mediating the exploration. This paper describes the first author's experience negotiating how Twitter data was mobilized from its site of production to the context of the AccLab-Mx, from the moment the use of social media data was decided to the reporting of insights from the analysis process.

A persistent feeling of uncertainty surrounded *the project*. Two factors contributed to such a feeling: For one thing, this was the first research project using social media data conducted by the AccLab-Mx. For another, nobody knew exactly how long would the pandemic last or how it would evolve. At the time of conducting the survey, Mexico City was on its highest alert due to COVID-19, and like many other cities in the world, it was in complete lockdown. The survey had been conducted online because, under such circumstances, only online-based forms of data collection were feasible. Following this logic, the idea of transitioning to social media platforms to continue the search for initiatives emerged in the discussions within the team. At the team meetings sessions, some expressed that both social and economic fatigue was probably the cause of the decline of citizen initiatives. As one of the partners said:

> *Social networks are a tool for massive diffusion of information, and in times of crisis they become a tool for organization. Now, they [social media platforms] are again a place to broadcast information.*

In Mexico, the predominant social media platforms for communication are Facebook, WhatsApp, and Twitter. However, current policies prevent collecting data from Facebook, and gathering data from WhatsApp requires belonging to a group on the platform. Informed of these limitations by the first author, the organization decided to run an exploratory analysis of what Twitter data could say about social capital and citizen capacity for action in Mexico City. Twitter, the first author explained to the AccLab-Mx, not only allows for data collection but is also the third most used social network in the country. According to recent statistics, there are eleven

[1] From here on we refer to the UNDP Accelerator Lab Mexico as AccLab-Mx.
[2] From here on we refer to it as *the project*.

[3] Alcaldía refers to a municipality. Mexico City is divided into sixteen municipalities, each of them under the control of a mayor (alcalde).

million Twitter users in Mexico, representing 60% of Internet users between 16 and 64 years old [14].

3.2 The Data Mobilization Process

Using the Twitter API, the fist author collected data between February 28th and May 17th, 2020 covering both the beginning of the pandemic—when imported cases where identified—and the community spread period [6]. The collection of tweets was filtered using the name of the sixteen *alcaldías* of Mexico City in combination with the hashtags and keywords collected in the preliminary analysis.

After collecting the data from Twitter, the first author implemented a natural language processing (NPL) tool for Spanish language to support the analysis of the collected data from Twitter. The input for the tool was a set of cleaned tweets from an *alcaldía* and a set of sentences describing a given citizen-led initiative. A second stage consisted of further reducing the number of tweets for qualitative analysis. Using the randomize tool in Excel, the team selected a sample of 100 tweets per alcaldía. Then, following the AccLab-Mx's definitions of "bonding", "bridging", and "linking", each tweet was categorized into one type of social capital. As a final step, the AccLab-Mx used content analysis to identify commonalities, distinctions, and relationships amongst the local responses that citizens, governments, and grassroots efforts were coordinating per *alcaldía* to address the challenges of the COVID-19 crisis. The emerging themes reflected the different manifestations of linking and bridging across *alcaldías*. Bonding could not be identified in this study. This paper's findings describe in detail the first author's experience facilitating this process for the AccLab-Mx and the challenges they—the first author and the AccLab-Mx—faced down the road.

3.3 The Mediator's Role

In this subsection, we provide a description of the nature of the first author's collaboration with the AccLab-Mx as a mediator between data and human resources and the organization itself. As an intern within the AccLab-Mx, collaborating on *the project*, the first author was in charge of the following tasks:

- **Assisting with the coordination of the volunteers:** During the three weeks in which the survey was conducted, the first author helped coordinate the volunteers that were in charge of distributing the survey.
- **Qualitative analysis of the survey's questions:** before shifting to collecting data from social media, the first author was asked to qualitatively analyze the survey answers to gain a deeper understanding of the information the AccLab-Mx was interested in collecting. As a result, the first author identified a set of characteristics of the citizen-led initiatives that were relevant for the AccLab-Mx. Based on these characteristics, a definition of what counts as a "citizen initiative" [4] was agreed. The initiatives were organized according to their purpose, which included food, health, education, labor, and public communication. As an example, Table 1 shows a sample of the attributes collected for initiatives that focused on food.
- **Data collection on Twitter:** based on the previously-agreed definition of "citizen-led initiative," the first author was assigned the collection of relevant Twitter data. She compiled a list of hashtags, keywords, and social media accounts of organizers of initiatives coordinated via social media. A total of 300,361 tweets were collected at the first stage.

[4]The AccLab-Mx concluded that relevant "citizen-led initiatives" are individual or collective actions that aim to address or reduce the negative impact of social, economic, or health problems caused by the COVID-19.

As part of the execution of her role as a mediator, the first author was included in internal communication (verbal and written), team meetings, and feedback rounds in which the specific definitions and types of social media data that were relevant to the AccLab-Mx and *the project* were negotiated. That position as first-hand witness and active participant in *the project* allowed the collection of ethnographic data that is at the core of the analysis that we present in this paper. In the next section, we delve into the characteristics of our method, the specific interactions and negotiations observed, and the analytical steps that we took to arrive at our findings.

4 METHOD

RQ: What are the implications of mediating how NGOs mobilize social media data from their site of production to inform institutional decisions for addressing humanitarian crises?

(1) (RQa) What are the practices that mediators engage in when supporting NGOs?
(2) (RQb) How do mediators and NGO teams negotiate what social media data counts as evidence for their work and what challenges do they encounter?

Following this research question and its sub-questions, we explore in this paper how the first author worked with the AccLab-Mx to define how to collect, interpret, and extract insights from social media data in the context of the COVID-19 pandemic and *the project* (see Section 3.1). The motivation for conducting this investigation is showing how, with the support of a mediating data expert, a large international organization explores social media data to understand the role of grassroots initiatives in the unique context of a Global pandemic. In particular, this research examines—from the perspective of the mediator—the process that the AccLab-Mx undergoes to decide on the use of social media data, collectively create ground-truth data, and negotiate their meaning, including the limitations encountered all along.

Our analysis is based on nine months of remote fieldwork conducted with the AccLab-Mx in the context of an internship held by the first author. This investigation is a practitioner-research project [15, 42, 62] in which we used semi-ethnographic methods such as participant observations, reflexive first-person accounts, and the analysis of text documents, especially fieldnotes and memos. To gain more immersive and participatory access to the negotiations involved in mobilizing social media data, the first author actively worked in *the project* for nine months between February and October 2020, serving as a research scientist working with four other team members in charge of collecting and analyzing data as instructed by the AccLab-Mx. During this time, the first author participated in meetings and feedback rounds, and engaged in email communication with different stakeholders and the research team. In total, her engagement comprised around 500 hours in which she assisted the AccLab-Mx in various capacities including note-taking, workshop facilitator, and the collection and analysis of the data used in *the project.*

4.1 Data Collection

The data collected consists primarily of interactions among team members participating in *the project*, between members of the AccLab-Mx and project collaborators, and also involving other stakeholders such as members of grassroots organizations. Those interactions were registered in detailed notes and memos. In addition, the first author registered reflexive accounts of her role, influence, and subjective interpretations about her fieldwork.

Those notes comprise thorough accounts of more than 30 meetings, all of them conducted online. Some of those meetings were held between the first author and one member of the AccLab-Mx, in which the first author would provide updates on her work. In other meetings, especially team meetings, the researcher on the field would take a rather passive role in which she would listen and take notes. Those meetings were oriented towards discussing *the project* and its main goal, that is, to understand existing capacities via various forms of analysis, including measuring three different forms of social capital—"bonding," "bridging," and "linking "— connecting communities and governments.

During those discussions, the first author observed the interactions and negotiations between different actors, that would stir *the project* in specific directions. In addition, feedback rounds involving the whole research team, other members of the AccLab-Mx, and partner grassroots organizations were held periodically. At those team meetings, next steps would be negotiated, decisions made, concepts defined, and phenomena interpreted. In total, we collected and analyzed 70 pages of field notes (including meeting minutes, observations, and self-reflexive accounts) and 47 documents (including instruction documents, email communication, and power point presentations).

4.2 Data Analysis

Mayring's qualitative content analysis was applied to analyze that data in view of the present investigation [85]. This approach aims at interpreting the manifest and latent content of the material in their social context and field of meaning, focusing on the personal perspective of the actors. This method allowed us enough flexibility to obtain valuable insights out of the personal notes and minutes taken by the first author. Some of the topics that had been operationalized in guidelines followed to conduct *the project* were indeed related to our research interest and helped us build deductive categories for the coding of the material. In addition, room for new categories was left open, so that they could be added after the exchange between recorded material and theoretical standpoints had taken place (inductive category formation). The development of coding schemes for the analysis, including categories and sub-categories, as well as the coding process itself was carried out in iterations involving cross-coding between all authors. We strove for interpretations that are intersubjectively comprehensible [61], exhaustive, and yet reflective of researchers' subjectivities.

4.3 Our Positionality

Our reflections on developing approaches to examine social media data to inform the work of non-profit organizations are shaped by our collective experience as researchers working with and investigating the production of data. Our decision to examine social media data re-use with a critical data studies perspective stems from our scholarly upbringings. We draw from the field of critical data studies, which argues for understanding data as situated and socially constructed rather than as an independent entity [79]. Perspectives from this field provide helpful lenses to outline the implications of effectively operationalizing the integration of social media data into institutional contexts to be used as evidence to inform community and policy interventions.

The first author, who is also the researcher in the field working with the AccLab-Mx, was born and raised in Mexico City, the place where this research takes place. She is familiar with the history of the city and the particularities of most of the neighborhoods. Although she is originally from Mexico, her postgraduate education has been in the Global North. At the time of this research, the first author was an intern at the AccLab-Mx in Mexico. The members of the AccLab-Mx were aware of her doctoral research and were supportive when she proposed to conduct the analysis that we present in this paper in parallel to the official study. In this sense, the first author considered herself as a mediator between the AccLab-Mx and the data work.

The second author is an Ecuadorian female professor of Human-Centered Computing at the Escuela Superior Politécnica del Litoral in Guayaquil, Ecuador. Her research, engaging in co-design with various vulnerable communities in the United States and Latin America is shaped by an assets-based approach; she works with communities in understanding and leveraging their strengths rather than trying to fix their deficits or lacks.

The third author is a female, Latin American HCI researcher working in a Global North institution at the intersection of sociology and computer science. Her work and her analysis of data and data mobilization is shaped by the specific epistemologies of these two disciplines — a relationship that is challenging, to say the least.

The fourth author is a male Latin American computer scientist working in the U.S. His role in this research was to design, implement, and evaluate the NLP algorithms used in *the project* to analyze the dataset.

The fifth author is a Caucasian male and associate professor of Human-Centered Computing in the U.S. His research is focused on participatory design approaches in partnership with underrepresented and under-served communities.

The role of the second, third, and fifth authors in relation to the investigation reported in this paper was to extrapolate and reinterpret data originated in a very specific context, that is, from the first author's reflections. In that sense, their task throughout this investigation has been to help the first author produce self-reflexive accounts from her fieldnotes, establish inter-subject comprehensibility when coding the data [61], and extract valuable high-level findings from the accounts reported by the first author.

5 FINDINGS

Our findings highlight the specific trade-offs, decisions, and assumptions that a non-profit organization might encounter when working with a mediating data expert to explore how to extract value from social media data. While these decisions and activities are rarely discussed by computational social scientists and social computing researchers analyzing digital traces to characterize and predict human behavior, they heavily influence the conclusions that can be drawn from social media data. Our results stress existing findings on how those decisions are bounded by the organizational needs and practices of the institutions guiding the work of data. Further, they provide evidence on the role of the data expert in using the AccLab-Mx information needs as input to inform concrete decisions that entailed defining, categorizing, or negotiating with data.

Our reflexive analysis draws attention to four key moments from *the project* in which decisions the first author made with the AccLab-Mx about how to collect and interpret data from Twitter informed *the project*. These four key moments are (1) defining the dataset, (2) defining ground truth, (3) making sense of data, and (4) extracting value of data and communicating insights. Each of these moments is based on the observations and experiences of the first author working within the AccLab-Mx and will be described in detail in the following subsections. To improve readability we will refer to the first author in her role as researcher involved in *the project* as "the researcher."

5.1 Moment 1: Defining the Dataset

The interpretation of data starts with the definition and collection of datasets, which are never given and are always defined in advance [4, 30, 47]. The analysis of Twitter data to identify up-to-date people's strategies for addressing the COVID-19 crisis across the neighborhoods—or alcaldías—of Mexico City began by defining the dataset of tweets to be analyzed. This process required a clear definition of keywords and time ranges that could illuminate or obscure patterns of interest. The AccLab-Mx had done work exploring citizen-led initiatives via surveys and had defined a clear need for connecting the behavior and capacities of those initiatives with each alcaldía in the city. Having been a participant during that previous data collection phase drove *the researcher* to deem it critical for the keywords and timeline defining the dataset to stem from those existing goals and efforts. In that regard, she proposed the following three considerations for constructing the dataset.

First, she decided to turn to the initial survey's responses for guidance on how to identify what tweets described citizen's initiatives. For running the surveys, the AccLab-Mx did not need to define what citizen's initiatives were; the citizens answering the survey could easily define them and provide information about them. When taking this query back to social media data, however, it became obvious that a definition was needed: an algorithm could not interpret what citizen initiatives were and decide how to recognize them on its own. The responses of one of the open-ended questions of the survey asking for a description of citizens' initiatives had information that could help craft this definition. *The researcher*, who had helped in the qualitative analysis of these responses, realized these could provide situated descriptions of how people characterize the strategies they observed and implemented in their communities to navigate the pandemic. She read 147 responses and compiled a list of keywords, hashtags, and social media accounts of people and organizations that coordinated initiatives. A more detailed description of the initial sample of keywords is in Table 1.

Second, *the researcher* proposed to prioritize the sixteen alcaldías of Mexico City in how the dataset was crafted by also using the names of the alcaldías as keywords to guide the data search. For the AccLab-Mx, it was imperative to gain a localized understanding of existing capacities, including social capital. Therefore, it was necessary to associate the content of the tweets with their location. However, one of the well-reported limitations of using social media data is determining the exact location of users who produce the content [45, 68, 74, 96]. While some techniques have been proposed to navigate this constraint [24, 76], *the researcher* decided to rather harness her knowledge of the city's geographical distribution and chose to filter the search of tweets using the names of alcaldías. Ensuring the tweets obtained did refer to alcaldías of Mexico City, however, required more than just using the names of alcaldías to filter the data search. Iterative manual analysis of small amounts of data related to citizen initiatives and collected using the first set of keywords, showed that people on Twitter tended to use acronyms and abbreviations when mentioning the name of their alcaldías. Consequently, different terms to refer to alcaldías of the city were incorporated into the search criteria for tweets. 2 describes the variations in the names of the neighborhoods in Mexico City.

Also, when applying the name of alcaldías as keywords, some tweets were not about the municipalities within Mexico City but other towns and cities across the country that share the same name (e.g., Venustiano Carranza [5]). These mixed results led *the researcher* to document and keep track of which alcaldías' names were more likely to return a larger number of tweets that corresponded to different places than the ones the team was documenting. Overall, *the researcher's* proposal to include the name of alcaldías as search criteria and process to maximize the accuracy of the results allowed for certain nuanced tweets, which would not have captured otherwise, to emerge. For example, we identified the communities of chinamperos [6] , who are exclusively located in the alcaldía of Xochimilco. During the pandemic, they organized via online channels to sell and ship their crops across Mexico City, creating supply networks with other alcaldías.

Third, *the researcher* recommended to the AccLab-Mx to carefully define the time frame to examine. This was critical due to the fleeting nature of citizen initiatives [63]: while there can be an intense presence of efforts at the beginning of a crisis, they tend to disappear as time goes by. The question for the AccLab-Mx, thus, was "what counts as a citizen-led initiative that is worth analyzing based on the time it took place?". After discussing it, members of the AccLab-Mx decided they were interested in understanding how people organized at the beginning of the pandemic until the government enforced the most strict lockdown. Therefore, *the researcher*

[5]There are four towns in the country called Venustiano Carranza, and each of them is in a different state.

[6]*Chinamperos* are farmers that use chinampas, which are an agroecosystem of pre-Hispanic origin, artificially built in areas of the lake of Xochimilco [111].

Category	Hashtags	Keywords	Text Description
Alimentos	#ComidaParaHeroes, #mercadoSolidario, #ConsumeLocal, #CanastaVerde	Frutas, verduras, vales, despensa,comida, alimentos,hortaliza, mercado, restaurante, fonda, productor, agricola, cocinar, gastronomico, agricultor, huacal, viveres.	*Caravana que acerca la venta de frutas y verduras a precio solidario a distintas colonias en Tlahuac.*
Food	#FoodForHeroes, #solidarityMarket, #consumeLocal, #GreenBasket	Fruits, vegetables, vouchers, pantry, food, food, vegetable, market, restaurant, inn, producer, agricultural, cook, gastronomic, farmer, huacal, groceries.	*Caravan that brings the sale of fruits and vegetables at a solidarity price to different neighborhoods in Tlahuac.*

Table 1: A sample of hashtags, keywords, and text description extracted from the initiatives collected through the survey. The first row shows the original text in Spanish, and the second row shows the translation in English.

Name of the neighborhood	Initial wording	Alternative abbreviations and acronyms
Gustavo A. Madero	Gustavo A. Madero	GAM, gamadero
Venustiano Carranza	Venustiano Carranza	V.Carranza
Benito Juarez	Benito Juarez	B.J

Table 2: The first column contains the original names of the neighborhoods. The second column includes the initial wording used in the first search. Lastly, the third column shows the abbreviations and acronyms collected on Twitter.

collected data from February 28 until May 17, 2020. These dates corresponded to the day when the Mexican health ministry detected the first imported cases of COVID until the day that community spread was declared, resulting in the lockdown [44, 52].

Applying these keywords and filters as criteria for searching data allowed for 300,361 tweets to be collected and used in the next phase.

5.2 Moment 2: Defining Ground Truth

Once the dataset to be analyzed was defined, the next goal for the AccLab-Mx was figuring out a way to examine what these tweets could say about citizen-led initiatives' capacities. In particular, the AccLab-Mx was interested in identifying the three different types of social capital that [9, 10] proposed to understand possibilities for action, locally (i.e., "bonding", "bridging", and "linking"g). It was, thus, critical for *the researcher* to help the AccLab-Mx identify canonical examples that could be associated with these forms of capital, leading to the definition of ground-truth. Ground-truth labeling of the data is a standard practice within the context of Machine Learning (ML) [21, 93, 102] that defines what a ML models should learn to mimic. Ground-truth data is created in a process that includes the ascription of specific meanings to data through labels which are typically based on the assumption that there is a single right answer for each data point and labelling instance [20].

While at this point of *the project*, *the researcher* had not yet decided what model to use for the analysis of tweets, she did know that she would use ML to analyze the dataset, making it necessary to define the ground truth. That is, what would make tweets to fall into the "bonding", "bridging", or "linking" classification. While the AccLab-Mx had given *the researcher* an explicit definition for each term, the representation of these phenomena in Twitter data was diverse. As such, the researcher had to engage in a negotiating process with members of the AccLab-Mx to co-develop more precise—but also situated—definitions together with canonical example of these behaviors.

To this end, *the researcher* held a calibration session with the AccLab-Mx team where she showed them a selection of tweets that she thought could encode the characteristics that they envisioned for each forms of social capital. To better illustrate how the negotiation and agreement around the ground-truth data took place, we describe in the following how the categories "bonding", "bridging", and "linking" were discussed and defined.

"Bonding" was defined at first by AccLab-Mx as "the connection among individuals that share similar characteristics." *The researcher* struggled to envision tweets that could fit this category; it sounded too broad and ambiguous for her and asked for clarification. The AccLab-Mx reframed their understanding of bonding as "the interaction of people that belong to equal or similar social groups, or that belong to the same community." The overall goal behind this definition was to determine the existing capacities of similar groups in coming together to support the same goal. *The researcher* searched for tweets that met this definition but only found one.

The tweet in question is shown in figure 2 [7] and describes how the families of incarcerated people organized donations of medicine to the prisons. Since the families organized the donations, *the researcher* inferred that aid was occurring between "similar groups," as required by the "bonding" definition. The AccLab-Mx agreed that this type of tweet reflected what they were looking for; that is, which groups of people were helping each other. Due to the difficulty of finding other examples in which the involvement of "similar groups" would be clear, however, *the researcher* suggested not to continue searching for this type of social capital and the AccLab-Mx agreed.

"Bridging" was defined as *the relationship between groups of people who seem to have less in common*. This concept could enable the AccLab-Mx to identify the types of groups that were in need for help and the ones that where able to provide help. As *the researcher* searched for exemplar tweets that could be classified as bridging, her main concern was how she should determine that the groups interacting were different. In the first calibration session, *the researcher* showed the AccLab-Mx team the tweets in figure 1 as she considered them to reflect a connection between different groups. However, members of the AccLab-Mx decided that only tweets similar to figure 1a and figure 1b should be classified as "bridging" for only those showed *a clear* difference between the group of people providing help and the group receiving aid. The tweet in figure 1c was not considered an appropriate example of "bridging" because it was about a business focusing on payments, package delivery, and food collection rather than voluntary assistance. Thus, it was established that "bridging" tweets should provide enough information to identify the group or individual offering help as well as the one receiving it, and that it is a voluntary donation and not a business.

"Linking" was defined by the AccLab-Mx as "a type of social capital that describes the formal and informal relationship between citizens and government authorities." This information could shed light on the different capacities that local authorities had and the citizens' abilities to engage with such capacities. Based on this definition, *the researcher* initially sought tweets that described programs and actions organized by government staff to serve the citizens and to reduce the impact of COVID-19 (see figure 3). For instance, for *the researcher*, the tweets in figure 3 classified as "linking" because they described government initiatives. Tweets related to different institutional actors also fit *the researchers'* understanding of linking (e.g., 3a published by the mayor of the *alcaldía* Iztapalapa, 3b published by the Mexican newspaper La Jornada, and 3c published by a citizen about the the *alcaldía* Iztapalapa). While the AccLab-Mx team agreed that these tweets represented "linking", one team member also raised the concern that these tweets might be too limiting in terms of the government-citizen interactions they showcased: "we do not want to put together a catalogue of all the government initiatives in response to the pandemic." The AccLab-Mx team then further reflected on their goal; they wanted to gain deep insight about governments' different levels of resourceful and creative behavior for supporting citizens during the crisis. This entailed COVID relief efforts and beyond. These reflections helped *the researcher* to look at the collected tweets from a different light; she no longer restricted her exploration to government's relief efforts but included any type of citizen-government interactions. Those tweets did offer more details of how varied *alcaldías'* ideas and capacities were. For example, while the mayor of alcaldía Xochimilco advertised the creation of WhatsApp groups to communicate directly with residents, the mayor of alcaldía Azcapotzalco, advertised YouTube broadcasts and videos about government programs updates. She now also considered tweets such as 4c, which highlighted that, in this alcaldía, existing capacities motivated citizens to report crimes to the authorities. Historically, this is not how citizens behave in Mexico; high levels of distrust towards authorities prevent crime reporting from taking place [5]. [8] In broadening the criteria for classifying tweets as "linking", thus, more interesting, promising local behaviors were able to emerge.

5.3 Moment 3: Making Sense of Data

Once the AccLab-Mx and *the researcher* defined what counted as ground truth, *the researcher* conducted a first iteration of analysis to explore the type of insights and outcomes that could be of interest to the AccLab-Mx. *The researcher* began by classifying some tweets according to their characteristics as "linking" or "bridging"; she did not categorize any tweet as "bonding" because it had been decided that category would be left out. While classifying the tweets, *the researcher* began reflecting on what could be the potential insights that this analysis could engender, and in particular, what of these insights could be of value to the AccLab-Mx.

The previous survey work the AccLab-Mx conducted suggested a desire for an exhaustive report of all relief initiatives, citizen- and government-led. After reviewing and classifying tweets, *the researcher* understood, however, that the observations she had collected from Twitter would never be able to provide such a report. While a Twitter data analysis could indeed identify initiatives, it was hard, if not impossible to ensure this data could really represent all existing initiatives. A richer finding, *the researcher* thought, would be a descriptive narrative of the collective organization across Mexico City; that is, nuanced evidence of local collective responses, as well as capacities, limitations, and other particularities of each alcaldía. Therefore, rather than quantifying citizens' initiatives, *the researcher* focused on identifying the communities' capacities that other research methods used by the AccLab-Mx (e.g., government statistics) could not uncover or explain. This approach is consistent with a perspective that understanding data is situated and socially constructed [45].

The decision of unpacking people's ongoing reactions to the pandemic from Twitter data entailed a shift of perspective for the AccLab-Mx. It was now needed to conduct another calibration meeting to give a new look at the previously collected tweets and decide *how* to distill full narratives about communities' capacities from them. When re-examining tweets describing "linking" social capital, one of the AccLab-Mx team members pointed out that, to really

[7] To maintain the anonymity of the citizens from whom we use tweets, we removed the user name and avatar.

[8] For the last ten years, the results of the National Survey of Victimization and Perception of Crime (In Spanish Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública - ENVIPE) have stated that more than 80% of crimes are not reported to the police. In fact, the last edition of the ENVIPE reported that in 2019 the population in Mexico suffered more than 30 million crimes, of which 92.4% were never reported to the authorities. According to the survey, victims decided not to report the crimes to the police because 36.3% consider it a waste of time, and 17.5% distrust the police.

(a) Tweet describing a restaurant's initiative that donated 150 dinners for medical staff

(b) Tweet from a citizen showing a group of neighbors donating food

(c) Tweet of an emerging business focused on helping people running their errands

Figure 1: Examples of the type of tweets *the researcher* associated with the category of bridging

Figure 2: Example of a tweet categorized as bonding

understand the impact of government authorities' tweets on citizens and communities, it was critical to pay attention to people's reactions. Specifically, he made a call to consider the number of likes and retweets, as well as people's comments to government's tweets. He argued that documenting "linking" social capital only from the perspective of government's posts was "like recording half of the conversation". For the AccLab-Mx team, it was becoming clearer that tweets entail two-way interactions between different actors, in this case, between government and citizens. To fully understand Twitter data, thus, they requested these interactions to be documented and analyzed. *The researcher*, then, recorded the likes and retweets along with the original tweets that she had classified as "linking," and added memos based on the tweets' comments.

In regards to the tweets classified as "bridging",members of the AccLab-Mx also discussed additional information to consider for gaining a more complete understanding of this social capital. They concluded that, for each "bridging" tweet, it was important to document the organizers and target populations as essential context-based information.

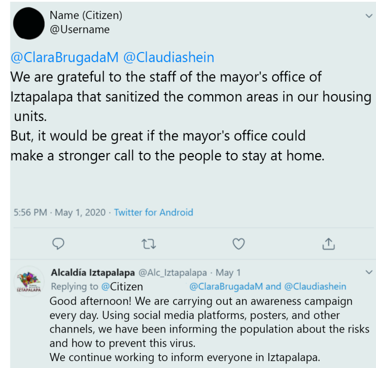
5.4 Moment 4: Extracting Value of Data and Communicating Insights

After defining the dataset (see 5.1), the ground truth (see 5.2), and the additional information needed to ensure the tweets could help generate valuable insights (see 5.3), *the researcher* used a combination of ML and qualitative methods to actually extract these insights from the dataset. At this stage, *the researcher* began to consider what would be the most appropriate way to communicate and present the findings to the AccLab-Mx. To that end, she struggled with two challenges. First, how to help the AccLab-Mx recognize that, in coming from social media data—that was taken out of its site of production and re-used for meeting the AccLab-Mx goals, the findings were coming from incomplete data. Second, how to organize the findings—which entailed a descriptive narrative rather than numbers and statistics—in ways that could be of value to the AccLab-Mx.

The first challenge entailed raising awareness to the AccLab-Mx about the nature of the findings that social media data can produce. *The researcher's* familiarity with literature discussing the limitations of representation within social media datasets made her keen to prevent the AccLab-Mx from seeing the findings as an objective, indisputable truth. Further, she knew the importance of making the AccLab-Mx aware of how their analytical path—from the definition of the dataset onwards—created a series of biases precisely feeding into those limitations. To promote a more critical approach to the AccLab-Mx understanding of the findings, she chose to start her presentation by explaining the incomplete nature of that data that produced these findings. She then used the memos about each decision behind the dataset construction that she recorded throughout the data analysis process to communicate the constraints of the data and the findings to the AccLab-Mx. In particular, these memos were useful to provide the AccLab-Mx with

(a) Tweet from the mayor of the Iztapalapa alcaldía describing the food program *Mercomuna*

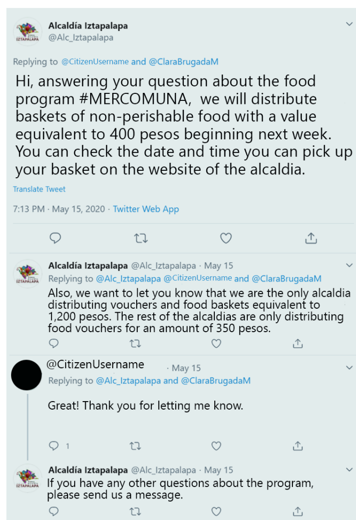
(b) Tweet from a newspaper with a video of cleaning squads sanitizing irregular settlements



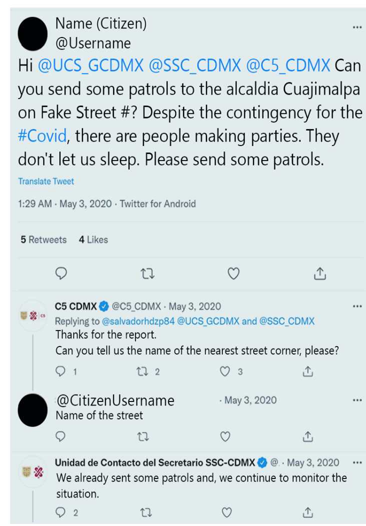
(c) Tweet from a citizen thanking the mayor's office for the sanitizing squads

Figure 3: Tweets showing targeted actions that the mayor's office of the *alcaldía* Iztapalapa undertook regarding food and public health.

(a) Tweet from the mayor of the alcaldía Iztacalco advertising the recording of a video conference on YouTube



(b) Conversation between a citizen and Twitter's official account of *alcaldía* Iztapalapa regarding a food program



(c) Citizen's report to the C5 (government video monitoring and emergency service) about people drinking alcohol on the streets

Figure 4: Examples of multiple engagement strategies between government staff and citizens.

examples of how specific decisions shaped not only the dataset but the insights that were derived from it.

The second challenge was one of information presentation and organization. While the AccLab-Mx and *the researcher* had agreed for the findings to be narrative-based, presenting such narratives so as to elicit insights that retained the context in which the social media data was originally produced, was a challenge [36, 45]. Knowing that the context of production would most likely get lost, throughout the process *the researcher* kept a log file recording the nuances behind each tweet. For example, in the log, each tweet was classified as either "linking" or "bridging" and each was recorded together with information about the initiatives the tweets described, including the target population, people's reactions (only for linking tweets), government-citizen communication methods (only for linking tweets), and initiative organizers (only for bridging tweets). When it came the time to present findings to the AccLab-Mx, *the* *researcher* went back to the log in search for guidance on how to organize such findings. She realized the tweets also provided rich information about the purpose of initiatives (e.g., securing water, covid testing, etc.) and the social context of each alcaldía, including local problems, assets, and existing infrastructures.

The researcher's experience with the AccLab-Mx, not only during the social media data analysis process but during the survey data collection, allowed her to realize this information also entailed a critical finding. For the AccLab-Mx, it was relevant to gain a granular and nuance view of how citizens, local governments, and non-profit organizations were organizing both on each alcaldía and in the city as a whole. The goal was for that view to spark ideas for collaborations and interventions in the city. She, thus, decided to present the findings dividing them in three categories: findings about social context, about "linking", and about "bridging". As shown in Figure 3, each category offered a summary of the information recorded in the log file. These categories highlighted which groups were helping each other, who was providing help, who was receiving it, and which were the pressing needs the non-profit sector was covering. The findings showed grassroots and nonprofit organizations were organizing most of the collective responses through donations.

6 DISCUSSION

Making sense of data to obtain value from it is a collective effort [103] constrained by tools [60, 110], disciplinary contexts [101, 106], organizational hierarchies, and power dynamics [92, 93]. Previous work in the fields of HCI and Critical Data Studies has developed recommendations and frameworks that draw from decolonial and feminist lenses to inform and guide practitioners in their perspectives when working with data [35, 43, 53, 97]. Similarly, various methodologies have been developed to uncover the multiple elements involved in the making and understanding of data [24, 76], thus, responding to the call of developing methods that provide a more situated and reflexive approach [35, 79, 95, 106, 107] rather than falling into the reductionist and technocratic epistemologies, which assume that data are objective and neutral [46, 79]. However, most of these previous efforts have focused on traditional types of datasets (e.g., government records, census data, etc.) and data collected by non-profits about their operations. Our investigation expands this line of inquiry with a set of reflections to guide data practitioners of non-profit organizations in their decision-making process when dealing with social media data.

Integrating social media data—which is often produced in highly particular contexts—into the work of non-profit organizations in the humanitarian space creates new demands on those involved in the value-discovery process of data. Specifically, the use of these data requires those conducting the analysis (e.g., data experts) as well as those consuming the findings (e.g., non-profit organizations and policy-making actors) to attain a rich, in-depth understanding of the potential advantages, constraints, and risks of using this type of data. The particular limitations of social media data—including limited representation of perspectives and the challenge of preserving context—demand organizations to be extra careful and strategic when drawing insights and translating them into humanitarian responses [25, 45, 104, 122]. As the analysis in this paper suggests, social media data limitations cannot simply be addressed through computational techniques.

Building on our analysis, we now unpack three assumptions that researchers using computational techniques to promote the use of social media data in non-profit organizations might face throughout their journey. To promote reflexivity upon each assumption, we provide in the next subsections a set of questions derived from our analysis. These questions are relevant for data experts seeking to support non-profit organizations. In doing so, we contribute to existing work proposing guidelines and frameworks for engaging in data work from a *situated* perspective. In line with Mohamed et al. [97], the questions we pose do not sum up to a conclusive method nor a prescriptive guideline. They are prompts for supporting data experts as they confront the following assumptions about the use of social media data in the humanitarian sector [53, 79]: *First*, that the use of social media data is straightforward, fast, and less expensive than using other types of data. *Second,* that the assumed objective nature of the findings is the main value that social media data provides to organizations. *Third*, that the effectiveness of the data expert's work is based on the computational methods they use and not on the situated knowledge of the data expert.

To summarize, the contribution of this research is the empirical evidence of the limitations that entail integrating social media data into the work of non-profit organizations and the questions we propose to data experts to prompt reflection in identifying and facing the limitations of this type of data. In the context of humanitarian work, these limitations require constant reflection and revision as they can have an immediate impact on communities' well-being and livelihood.

6.1 Assumption 1: The use of social media data is straightforward, fast, cost-effective

Increasingly, a range of actors in the field of humanitarian action are proposing the analysis of social media data as a promising pathway for non-profit organizations [54, 87, 89]. Various factors could lead organizations to rush into this option. By conveying a sense of being abundant, openly available, and entailing a deluge of different experiences and opinions [78, 80], social media data tends to be portrayed as an efficient and effective solution to inform organizations' operations for *1)* knowing how to deploy their resources [8]; and *2)* reporting impact back to funders [40, 80, 89]. Further, and perhaps more concerning, organizations can perceive the use of social media data as an easy and quick step towards technological innovation, which has increasingly been presented as a need for those deploying humanitarian support on the ground [69, 89].

However, our analysis of the first author's experience clarifies that distilling value from social media data is not a quick nor cost-effective solution. As we saw, the abstraction of value from social media demands at least three types of resources from non-profit organizations: time, flexibility, and freedom. The process was not only time-intensive, demanding various calibrating sessions for reviewing the nature, potential, and implications of different tweets, but it also required much expertise from the mediator to guide the team during those sessions. As such, a situated social media data analysis entails laborious tasks similar to those demanded by qualitative data analysis as Maitland et al. [81] reported. More importantly,

Category	Subcategories	Description
Social Context: Tweets that refer to social problems of alcaldias	Infrastructure	Refers to the public resources of the alcaldias
	Local Problems	Situated issues
	Community Assets	Local capacities
Linking: Ties between communities with local government	Type of initiatives	Documented and categorized the purpose of initiatives
	Target Population	Documented and categorized the groups of people that benefit from government's initiative
	People's reaction	Registered number of tweets, likes and comments
		Classified comments into three categories: complaint, approval, other
	Communication between government and people	Make memos describing the methods each alcaldia used to communicate with the people
Bridging: Connections across communities	Type of initiatives	Documented and categorized the purpose of initiatives
	Target Population	Documented and categorized the groups of people that benefit from the initiative
	Organizers	communities and organizations that coordinated the initiatives

Table 3: Categories of analysis used to inform *the organization*

as the analysis shows, the process required willingness to revise and flexibility to fail. For example, when *the researcher* expressed confusion about the definition of "bridging" that the AccLab-Mx collaborators were abiding by, they were open to reconsidering it. Further, when not enough tweets emerged depicting the concept of "bonding", the AccLab-Mx was able to admit not using social media data to find evidence about that type of social capital. Lastly, the AccLab-Mx had the freedom to accept that the results would no longer be the numbers and statistics it was looking for and was open to accepting narratives instead.

Due to the nature of this particular organization—large, transnational, and resourceful—time and expertise were not significant constraints. Moreover, the raison d'être of the division conducting *the project* within the AccLab-Mx also afforded it the freedom to fail; it was created precisely to explore innovative solutions to humanitarian relief problems. As such, its use of social media data was only one option amongst many that were being explored simultaneously by other teams. For most organizations in the area of humanitarian action, especially the small ones working close to communities, time, expertise, willingness to revise, and flexibility to fail are extremely scarce resources. Most non-profits engage in highly informal operations, led by creativity and need rather than by data analysis expertise [27, 56, 73, 109]. If they happen to engage with data experts, these are not necessarily part of their everyday staff [56, 72] and thus, cannot really walk the organization through the exploratory process that social media data requires. Further, these organizations cannot afford either time to explore or failure: they often must act as quickly as possible to respond to urgent needs they witness on an everyday basis. While social media data analysis demands particular resources, our work suggests it is not a quick solution for the labor demands that data analytics poses to organizations [101]. Considering these resources and workloads demands, thus, should not be dismissed when deciding what data non-profit organizations can use.

Our analysis highlights that before recommending organizations to explore social media data, data experts and researchers need to carefully guide them in a self-assessment process. To do so, they can ask themselves and the organization *"what human resources does the organization have?," "what level of pressure to act is the organization facing?," "what exploration spaces can it afford?," and "in which cases can it revise or reconsider the goals it is trying to pursue?."*

6.2 Assumption 2: Objective findings are the main value that social media data provides

Organizations in general engage in datafication processes intending to obtain valuable results [26, 98]. For non-profit organizations working in humanitarian contexts, valuable results from datafication entail evidence for deciding when and where to act (e.g., the number and type of resources to mobilize and the communities that are in most need) as well as for reporting results to funders [26, 29, 34, 98]. Thus, what these organizations expect to gain from the data is tied to an idea of an objective truth that can guide their actions [89, 127].

Previous work has strongly contested the idea that the value of datafication can lie in the objective identification of meanings and patterns from data [33], helping measure or giving visibility to problems [67]. The notion of objectivity in data value disregards the intrinsic incompleteness of big data —and thus, the uncertainties it raises when trying to use it for making decisions—as well as the difficulties that organization's stakeholders face to negotiate what is of value to them [110, 120, 127]. Indeed, data analysis is a work of articulation plagued by moments in which human decisions and interpretation remove the possibility of obtaining "objective" results that represent absolute truth [46, 79, 81].

Instead, academics increasingly advocate for understanding the value of data in terms of the incomplete and imperfect stories it can tell about the research context, including the relations that it can highlight [26, 33]. Our findings extend that work by highlighting that, in the context of non-profit organizations using social media data to inform humanitarian work, such valuable, imperfect stories do not emerge as an outcome of data analysis but lie within the data analysis process.

The value abstracted from the project described in this paper, for example, lies in the narratives that the *researcher* presented, which revealed other actors and dimensions of the examined problem that the AccLab-Mx had not considered before. Moreover, there was also value in the revisions and reflections that the analysis process prompted. Every time *the researcher* and *the AccLab-Mx team* faced situations that demanded them to rethink their beliefs and definitions (e.g., what does it mean for two citizen-led initiatives to "belong to equal or similar social groups"?, what nuances the concept of linking should consider?, what to extract from Twitter if not the number of all citizen-led initiatives in the city?) they gained a more situated understanding of their assumptions, the context in the city of Mexico, and what Twitter could and could not offer them. All these forms of understanding, which helped the organization to reframe what data is for, added important value to the AccLab-Mx [43].

Our findings confirm what previous research has pointed out in terms of challenges when extracting value or actionable insights from data [81]. The insights we present in this work contest the notion that data-related work is neutral and instead demonstrate that abstracting actionable knowledge from social media data is a process contingent on the organizations' needs and requires acknowledging the particularities of the data. For organizations that deliver humanitarian relief, integrating social media data into their work entails recognizing that data alone cannot fully reflect conditions "on the ground" [26, 33, 38]. Instead, organizations should approach user-generated content as a starting point that needs to be completed with additional information using other sources of data and methods.

When supporting the analysis of data for organizations, research has reported that the data expert can greatly contribute to organizational memory and knowledge-sharing [56]. As our findings show, in the context of mobilizing social media data for non-profits and grassroots organizations, the data expert has yet another critical role: that of helping the organization recognize and appreciate these unexpected forms of value, to "reframe what data is for", which Couldry et al., recommend for decolonizing data [43]. To that end, it becomes important for the data expert to continuously ask themselves: *"what are the concepts, frameworks, and beliefs about human behavior that the organization has?,"* and notice when breakdowns in the analysis forces it to revise them. Data experts can also ask *"what are the types of decisions that the organization plans to inform through social media data analysis?"* and then, engage the organization in discussions about how changes in concepts and frameworks might entail a change in the feasibility of attaining their end goal. As we saw, rather than failures, those moments are opportunities for organizations to not only redefine what they seek and need but to grow a critical understanding of how and when to resort to social media data for adding value to their operations.

6.3 Assumption 3: The effectiveness of the data expert's work lies in the computational methods they use

A widespread assumption in work based on social media data is that its effectiveness lies in the methods used for the analysis, especially if the analysis is based on computational techniques [45, 55]. Loukissas [79] challenges such a belief by encouraging data practitioners to develop methods that consider the context where data is generated as part of the data analysis. This can help bring a more meaningful and responsible interpretation and use of data. Our findings show that, in the context of social media data, Loukissas' call remains relevant: *the researcher*'s commitment to developing contextualized methods had a critical role in the process of mobilizing Twitter data from their site of production to the context and goals of the AccLab-Mx. *The researcher* made sure that the AccLab-Mx would choose a social media platform that was relevant for Mexican citizens, proposed keywords such as the names of alcaldías, that would emphasize data's geographical context, and iterated on the use of the keywords to ensure they would provide as much rich information about alcaldías as possible.

In this sense, it is critical that data experts continuously reflect on their positionality and how it affects their data analysis abilities and decisions. This includes clearly identifying from the very beginning where they stand with respect to the type of results that can be drawn from the data. It also includes knowing the advantages and limitations of their cultural background, and of their level of understanding not only of the organization's goal but of the organization itself. This form of positionality is key to developing much-needed contextualized methods to mobilize social media data.

In the case of *the researcher*, she knew from the beginning of her work relationship with this initiative's team to a critical understanding of social media data's capacities and limitations. Her familiarity with the Critical Data Studies literature gave her a rich understanding of the negative impact that portraying social media data as a complete, objective solution could have for the AccLab-Mx [45, 53]. To promote a situated and critical perspective on the implications of using Twitter data, thus, she documented all the decisions of the process, emphasizing those that implied a loss of context or leaving aside nuances that could not be recovered later. Researchers engaging in mediating social media data work, thus, need to ask themselves *"What are my expectations in working with a non-profit organization?," "What potential negative effects do I foresee in introducing social media data practices to the organization?", "What do I feel is my responsibility in the way the organization perceives the potential of social media data?"*

Reflecting on one's cultural background, as our findings show, also plays a critical role when approaching social media data from a local perspective. It remains critical to recognize the role of our *lived experiences* in our ability to interrogate how data contributes to "the current configuration of structural privilege and structural oppression" [53]. Data experts who are not familiar with the context they are examining need to reflect on what they might be missing and how that might impact the organization. In the case of *the researcher*, it was not only her formal understanding of Mexico City but also the informal one gained after growing up and living there for twenty years, which led her to recognize abbreviations of the alcaldías' names in the tweets she reviewed. Furthermore, she noticed when certain tweets were about different places from those the AccLab-Mx was interested in examining. If she had not had this knowledge, all this information about citizens' initiatives would simply have been lost. *The researcher's* approach illustrates what Loukissas refers to as *reading data in place*, i.e., seeing data within an interpretive context [79]. Approaching data from a local perspective can help to understand their limitations and biases. Neglecting the local specificity of data might hinder finding appropriate forms of social advocacy [79]. When the data expert is not familiar with the context, they need to ask themselves *"What are the potential problems that can arise if I try to interpret language, culture, and beliefs that I am not familiar with?."*

Finally, our findings suggest that to ensure data is contextualized, a data expert needs to be able to always connect the data analysis process with the goals and work that their organization undertook before. Although the relationship of *the researcher* with the AccLab-Mx was temporary, she had been involved in many other data collection tasks before working with them on the social media data project. She, thus, understood not only the goal of the AccLab-Mx but what such an initiative had been doing and was doing in parallel to the work she was leading. This allowed *the researcher* to suggest alcaldías as keywords to filter, to use survey responses for defining what a citizen-led initiative is, and how to organize information in the presentation of the results. Before recommending pathways of action with data, it becomes, thus, essential for data experts to realize these recommendations cannot happen in a void and ask themselves *"what decisions have the organization already made?," "what data has it previously collected and why?," and "how do the organization' current goals relate to previous work and goals?".*

7 CONCLUSION

In this paper, we reported on nine months of fieldwork collaboration with the Accelerator Lab Mexico mobilizing social media data to define, identify, and categorize grassroots citizens initiatives aimed at addressing the social and economic challenges of the COVID-19 health crises in Mexico. Our findings showed four moments of negotiation that shape the mobilization of social media data as well as the insights that data is able to provide: (1) the definition of the dataset, (2) the establishment of the ground truth, (3) the sensemaking of the data, and (4) the extraction of insights and value. Those moments are signed by the goals and capabilities of humanitarian organizations, as well as the on-the-ground knowledge and positionality of the researcher mediating between NGO and data. In light of these findings, we discussed and challenged common assumptions about the use of social media data as being cheap, straightforward, and fast, and provided valuable lessons for data experts collaborating with non-profit organizations to mobilize social media data, extract insights, and, ultimately, inform paths for action in crisis-affected communities. We recognize that the findings we present are limited by the uniqueness of the COVID-19 crisis, the characteristics of Mexico City, and the questions the organization aimed to inform with the Twitter analysis. Ultimately, these conditions defined the methods and decisions the first author followed. Despite these limitations, the reflexive questions we propose in this work aim to be reflexive rather than prescriptive in recognition of the fact that engaging with the locality of data prevents strong moves toward generalization.

ACKNOWLEDGMENTS

This research partially used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

The research reported in this article builds upon the previous work completed by the first author as part of her Ph.D. at the Georgia Institute of Technology and her internship position with the Accelerator Lab at the United Nations Development Programme (UNDP) in Mexico. The findings, analysis, discussion, and recommendations of this research article, do not represent the official position of UNDP or of any of the UN Member States that are part of its Executive Board. They are also not necessarily endorsed by those mentioned in the acknowledgments or cited. The published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the UNDP be liable for damages arising from its use. We thank the team of the Accelerator Lab at UNDP Mexico for allowing us to build this research upon the COVID-19 Social Inventory project. Specifically, we are deeply grateful for the support of Jorge Munguía, Gabriela Ríos, Luis Cervantes, and Alicia López.

REFERENCES

[1] [n. d.]. Nuevo coronavirus 2019. https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019
[2] [n. d.]. Semáforo. https://coronavirus.gob.mx/semaforo/
[3] [n. d.]. United Nations Development Programme. https://www.undp.org/
[4] 2013. *Raw data is an oxymoron.* The MIT Press.
[5] 2019. Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2020. https://www.inegi.org.mx/contenidos/programas/envipe/2020/doc/envipe2020_presentacion_nacional.pdf
[6] 2020. Coronavirus en México. Confirman primera muerte por Covid-19. https://www.eluniversal.com.mx/nacion/coronavirus-secretaria-de-salud-confirma-la-primera-muerte-por-covid-19-en-mexico
[7] Mark S. Ackerman, Volker Wulf, and Volkmar Pipek. 2002. *Sharing Expertise: Beyond Knowledge Management.* MIT Press, Cambridge, MA, USA.
[8] Emily Aiken, Suzanne Bellue, Dean Karlan, Christopher R Udry, and Joshua Blumenstock. 2021. *Machine Learning and Mobile Phone Data Can Improve the*

Targeting of Humanitarian Assistance. Working Paper 29070. National Bureau of Economic Research. https://doi.org/10.3386/w29070

[9] Daniel P Aldrich. 2012. Social, not physical, infrastructure: the critical role of civil society after the 1923 Tokyo earthquake. *Disasters* 36, 3 (2012), 398–419.

[10] Daniel P. Aldrich and Michelle A. Meyer. 2015. Social Capital and Community Resilience. *American Behavioral Scientist* 59, 2 (2015), 254–269. https://doi.org/10.1177/0002764214550299 arXiv:https://doi.org/10.1177/0002764214550299

[11] Adriana Alvarado Garcia, Matthew J. Britton, Dhairya Manish Doshi, Munmun De Choudhury, and Christopher A. Le Dantec. 2021. Data Migrations: Exploring the Use of Social Media Data as Evidence for Human Rights Advocacy. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 268 (Jan. 2021), 25 pages. https://doi.org/10.1145/3434177

[12] Adriana Alvarado Garcia and Christopher A. Le Dantec. 2018. Quotidian Report: Grassroots Data Practices to Address Public Safety. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 17 (Nov. 2018), 18 pages. https://doi.org/10.1145/3274286

[13] Adriana Alvarado Garcia, Alyson L. Young, and Lynn Dombrowski. 2017. On Making Data Actionable: How Activists Use Imperfect Data to Foster Social Change for Human Rights Violations in Mexico. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 19 (Dec. 2017), 19 pages. https://doi.org/10.1145/3134654

[14] Clay Alvino. 2021. Estadísticas de la situación digital de México en el 2020-2021. https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-mexico-en-el-2020-2021/

[15] Gary L. Anderson. 1994. *Studying your own school : an educator's guide to qualitative practitioner research /*. Corwin Press, Thousand Oaks, Calif.

[16] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. 2022. *Human-centered data science: An introduction*. The MIT Press.

[17] Jay Aronson. 2018. Challenging Impunity with User Generated Content. https://www.ijmonitor.org/2015/06/challenging-impunity-with-user-generated-content/

[18] Jay D. Aronson. 2014. Crowd-Training Machine Learning Systems for Human Rights Abuse Documentation. In *AAAI 2014*.

[19] Jay D. Aronson. 2018. *The Utility of User-Generated Content in Human Rights Investigations*. Cambridge University Press, 129–148. https://doi.org/10.1017/9781316838952.006

[20] Lora Aroyo and Chris Welty. 2013. Crowd Truth: Harnessing disagreement in crowdsourcing a relation extraction gold standard. https://doi.org/10.6084/M9.FIGSHARE.679997.V1

[21] Lora Aroyo and Chris Welty. 2015. Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation. *AI Magazine* 36, 1 (Mar. 2015), 15–24. https://doi.org/10.1609/aimag.v36i1.2564

[22] Shrey Bagroy, Ponnurangam Kumaraguru, and Munmun De Choudhury. 2017. *A Social Media Based Index of Mental Well-Being in College Campuses*. Association for Computing Machinery, New York, NY, USA, 1634–1646. https://doi.org/10.1145/3025453.3025909

[23] Liam J. Bannon and Kari Kuutti. 1996. Shifting Perspectives on Organizational Memory: From Storage to Active Remembering. In *Proceedings of the 29th Hawaii International Conference on System Sciences Volume 3: Collaboration Systems and Technology (HICSS '96)*. IEEE Computer Society, USA, 156.

[24] Jo Bates, Yu-Wei Lin, and Paula Goodale. 2016. Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society* 3, 2 (2016), 2053951716654502. https://doi.org/10.1177/2053951716654502 arXiv:https://doi.org/10.1177/2053951716654502

[25] N. Baym. 2013. Data not seen: The uses and shortcomings of social media metrics. *First Monday* 18 (2013).

[26] David Bell, Mark Lycett, Alaa Marshan, and Asmat Monaghan. 2021. Exploring future challenges for big data in the humanitarian domain. *Journal of Business Research* 131 (2021), 453–468. https://doi.org/10.1016/j.jbusres.2020.09.035

[27] Lehn M. Benjamin, Amy Voida, and Chris Bopp. 2018. Policy fields, data systems, and the performance of nonprofit human service organizations. *Human Service Organizations: Management, Leadership & Governance* 42, 2 (2018), 185–204. https://doi.org/10.1080/23303131.2017.1422072 arXiv:https://doi.org/10.1080/23303131.2017.1422072

[28] Chris Bopp. 2019. Doing "Good" with Data? Understanding and Working Around Data Doubles in Human Services Organizations. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 33–37. https://doi.org/10.1145/3311957.3361850

[29] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3608–3619. https://doi.org/10.1145/3025453.3025694

[30] Christine L. Borgman. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.

[31] Geoffrey C. Bowker. 2000. Biodiversity Datadiversity. *Social Studies of Science* 30, 5 (Oct. 2000), 643–683. https://doi.org/10.1177/030631200030005001

[32] Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.

[33] Ryan Burns. 2018. Datafying Disaster: Institutional Framings of Data Production Following Superstorm Sandy. *Annals of the American Association of Geographers* 108, 2 (2018), 569–578. https://doi.org/10.1080/24694452.2017.1402673 arXiv:https://doi.org/10.1080/24694452.2017.1402673

[34] Ryan Burns. 2019. New Frontiers of Philanthro-capitalism: Digital Technologies and Humanitarianism. *Antipode* 51, 4 (2019), 1101–1122. https://doi.org/10.1111/anti.12534 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/anti.12534

[35] Isabelle Vonèche Cardia, Adrian Holzer, Ying Xu, Carleen Maitland, and Denis Gillet. 2017. Towards a Principled Approach to Humanitarian Information and Communication Technology. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development* (Lahore, Pakistan) *(ICTD '17)*. Association for Computing Machinery, New York, NY, USA, Article 23, 5 pages. https://doi.org/10.1145/3136560.3136588

[36] Samuelle Carlson and Ben Anderson. 2007. What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication* 12, 2 (01 2007), 635–651. https://doi.org/10.1111/j.1083-6101.2007.00342.x arXiv:https://academic.oup.com/jcmc/article-pdf/12/2/635/22317230/jjcmcom0635.pdf

[37] Joanne G Carman and Kimberly A Fredericks. 2010. Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation* 31, 1 (2010), 84–104.

[38] Sarah Carnochan, Mark Samples, Michael Myers, and Michael J. Austin. 2014. Performance Measurement Challenges in Nonprofit Human Service Organizations. *Nonprofit and Voluntary Sector Quarterly* 43, 6 (2014), 1014–1032. https://doi.org/10.1177/0899764013508009 arXiv:https://doi.org/10.1177/0899764013508009

[39] Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1171–1184.

[40] Dmitry Chernobrov. 2018. Digital Volunteer Networks and Humanitarian Crisis Reporting. *Digital Journalism* 6, 7 (2018), 928–944. https://doi.org/10.1080/21670811.2018.1462666 arXiv:https://doi.org/10.1080/21670811.2018.1462666

[41] Danielle Citron and Frank Pasquale. 2014. The scored society: Due process for automated predictions. *Washington Law Review* 89 (03 2014), 1–33.

[42] Marilyn Cochran-Smith and Lytle Susan L. 1993. *Inside/outside : teacher research and knowledge /*. Teachers College Press, New York.

[43] Nick Couldry and Ulises A Mejias. 2019. Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20, 4 (2019), 336–349.

[44] COVID-19. 2021. Se mantienen Semáforo Rojo y suspensión de actividades no esenciales en la Ciudad de México. https://covid19.cdmx.gob.mx/comunicacion/nota/se-mantienen-semaforo-rojo-y-suspension-de-actividades-no-esenciales-en-la-ciudad-de-mexico

[45] danah boyd and Kate Crawford. 2012. Critical Questions For Big Data. *Information, Communication & Society* 15, 5 (2012), 662–679. https://doi.org/10.1080/1369118X.2012.678878 arXiv:https://doi.org/10.1080/1369118X.2012.678878

[46] Lorraine Daston and Peter Galison. 2018. *Objectivity*. Zone Books.

[47] Hannah Davis. 2020. A Dataset Is a Worldview. https://towardsdatascience.com/a-dataset-is-a-worldview-5328216dd44d.

[48] Munmun De Choudhury, Scott Counts, Eric J. Horvitz, and Aaron Hoff. 2014. Characterizing and Predicting Postpartum Depression from Shared Facebook Data *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 626–638. https://doi.org/10.1145/2531602.2531675

[49] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2098–2110. https://doi.org/10.1145/2858036.2858207

[50] Munmun De Choudhury, Sanket S. Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 353–369. https://doi.org/10.1145/2998181.2998220

[51] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society* 8, 2 (July 2021), 20539517211035955. https://doi.org/10.1177/20539517211035955 Publisher: SAGE Publications Ltd.

[52] Milenio Digital. 2020. https://www.milenio.com/ciencia-y-salud/asi-fue-la-primera-muerte-por-covid-19-en-mexico

[53] Catherine DIgnazio and Lauren F. Klein. 2020. *Data Feminism*. MIT PRESS LTD.

[54] Kathrin Eismann, Oliver Posegga, and Kai Fischbach. 2018. Decision making in emergency management: The role of social media. (2018).

[55] M. C. Elish and danah boyd. 2018. Situating methods in the magic of Big Data and AI. *Communication Monographs* 85, 1 (2018), 57–80. https://doi.org/10.1080/03637751.2017.1375130 arXiv:https://doi.org/10.1080/03637751.2017.1375130

[56] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1273–1283. https://doi.org/10.1145/2818048.2820068

[57] Virginia Eubanks. 2019. *Automating inequality: how high tools profile, police, and punish the poor.* Picador.

[58] Hayley I. Evans, Marisol Wong-Villacres, Daniel Castro, Eric Gilbert, Rosa I. Arriaga, Michaelanne Dye, and Amy Bruckman. 2018. *Facebook in Venezuela: Understanding Solidarity Economies in Low-Trust Environments*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173802

[59] C. Featherstone. 2013. Identifying vehicle descriptions in microblogging text with the aim of reducing or predicting crime. In *2013 International Conference on Adaptive Science and Technology*. 1–8. https://doi.org/10.1109/ICASTech.2013.6707494

[60] Melanie Feinberg. 2017. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2952–2963. https://doi.org/10.1145/3025453.3025837

[61] Uwe Flick, Ernst von Kardorff, and Ines Steinke (Eds.). 2004. *A companion to qualitative research*. Sage Publications, London ; Thousand Oaks, Calif.

[62] Eric Gutstein. 2003. Teaching and Learning Mathematics for Social Justice in an Urban, Latino School. *Journal for Research in Mathematics Education* 34, 1 (2003), 37–73. http://www.jstor.org/stable/30034699

[63] Ira Helsloot and Arnout Ruitenberg. 2004. Citizen response to disasters: a survey of literature and some practical implications. *Journal of contingencies and crisis management* 12, 3 (2004), 98–111.

[64] Marco Iansiti. 2021. The value of data and its impact on competition. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3890387

[65] Amnesty International. 2011. Deadly detention:Deaths in custody amid popular protest in Syria. https://www.amnestyusa.org/wp-content/uploads/2017/04/mde240352011en_5.pdf

[66] Amnesty International. 2013. How is Social Media Transforming Human Rights Monitoring? https://tinyurl.com/y4wbee2o

[67] Azra Ismail and Neha Kumar. 2021. AI in global health: the view from the front lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–21.

[68] Isaac Johnson, Connor McMahon, Johannes Schöning, and Brent Hecht. 2017. The Effect of Population and "Structural" Biases on Social Media-Based Algorithms: A Case Study in Geolocation Inference Across the Urban-Rural Spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1167–1178. https://doi.org/10.1145/3025453.3026015

[69] Nayomi Kankanamge, Tan Yigitcanlar, and Ashantha Goonetilleke. 2021. Public perceptions on artificial intelligence driven disaster management: Evidence from Sydney, Melbourne and Brisbane. *Telematics and Informatics* 65 (2021), 101729. https://doi.org/10.1016/j.tele.2021.101729

[70] Naveena Karusala, Jennifer Wilson, Phebe Vayanos, and Eric Rice. 2019. Street-Level Realities of Data Practices in Homeless Services Provision. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 184 (nov 2019), 23 pages. https://doi.org/10.1145/3359286

[71] Marc-André Kaufhold, Christian Reuter, and Thomas Ludwig. 2019. Cross-Media Usage of Social Big Data for Emergency Services and Volunteer Communities: Approaches, Development and Challenges of Multi-Platform Social Media Services. *arXiv preprint arXiv:1907.07725* (2019).

[72] Vera Khovanskaya and Phoebe Sengers. 2019. Data Rhetoric and Uneasy Alliances: Data Advocacy in US Labor History. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1391–1403. https://doi.org/10.1145/3322276.3323691

[73] Vera Khovanskaya, Phoebe Sengers, and Lynn Dombrowski. 2020. Bottom-Up Organizing with Tools from On High: Understanding the Data Practices of Labor Organizers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376185

[74] Annice E. Kim, Heather M. Hansen, Joe Murphy, Ashley K. Richards, Jennifer Duke, and Jane A. Allen. 2013. Methodological considerations in analyzing Twitter data. *Journal of the National Cancer Institute. Monographs* (Dec 2013). https://pubmed.ncbi.nlm.nih.gov/24395983/

[75] Marina Kogan, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, and Stuart Geiger. 2020. Mapping Out Human-Centered Data Science: Methods, Approaches, and Best Practices. In *Companion of the 2020 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '20)*. Association for Computing Machinery, New York, NY, USA, 151–156. https://doi.org/10.1145/3323994.3369898

[76] Guillaume Latzko-Toth, Claudine Bonneau, and Mlanie Millette. [n. d.]. Chapter 13: Small Data, Thick Data: Thickening Strategies for Trace-based Social Media Research. http://methods.sagepub.com/book/the-sage-handbook-of-social-media-research-methods/i1797.xml

[77] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S Hopkins, and Nina Kruschwitz. 2011. Big data, analytics and the path from insights to value. *MIT sloan management review* 52, 2 (2011), 21–32.

[78] Brooke Fisher Liu, Lucinda Austin, and Yan Jin. 2011. How publics respond to crisis communication strategies: The interplay of information form and source. *Public relations review* 37, 4 (2011), 345–353.

[79] Yanni Alexander Loukissas. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society*. The MIT Press.

[80] Mirca Madianou. 2019. Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises. *Social Media + Society* 5, 3 (2019), 2056305119863146. https://doi.org/10.1177/2056305119863146 arXiv:https://doi.org/10.1177/2056305119863146

[81] Carleen Maitland, Jean-Laurent Martin, Maria Gabriela Urgiles Bravo, and Alex Bertram. 2022. A Qualitative Difference: Integrating Qualitative Data into Humanitarian Response Operations. In *Proceedings of the Twelfth International Conference on Information and Communication Technologies and Development* (Seattle, USA) *(ICTD '22)*. Association for Computing Machinery, New York, NY, USA, 25 pages.

[82] Vukosi Marivate and Pelonomi Moiloa. 2016. Catching crime: Detection of public safety incidents using social media. In *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*. IEEE, 1–5.

[83] Aaron Martin, Gargi Sharma, Siddharth Peter de Souza, Linnet Taylor, Boudewijn van Eerd, Sean Martin McDonald, Massimo Marelli, Margie Cheesman, Stephan Scheel, and Huub Dijstelbloem. 2022. Digitisation and Sovereignty in Humanitarian Space: Technologies, Territories and Tensions. *Geopolitics* 0, 0 (2022), 1–36. https://doi.org/10.1080/14650045.2022.2047468 arXiv:https://doi.org/10.1080/14650045.2022.2047468

[84] Nan L. Maxwell, Dana Rotz, and Christina Garcia. 2016. Data and Decision Making: Same Organization, Different Perceptions; Different Organizations, Different Perceptions. *American Journal of Evaluation* 37, 4 (2016), 463–485. https://doi.org/10.1177/1098214015623634 arXiv:https://doi.org/10.1177/1098214015623634

[85] Philipp Mayring. 2015. *Qualitative Content Analysis: Theoretical Background and Procedures*. Springer Netherlands, Dordrecht, 365–380. https://doi.org/10.1007/978-94-017-9181-6_13

[86] Lori McCay-Peet and Anabel Quan-Haase. 2016. What is Social Media and What Questions Can Social Media Research Help Us Answer.

[87] Anthony McCosker, Peter Kamstra, Tracy De Cotta, Jane Farmer, Frances Shaw, Zoe Teh, and Arezou Soltani Panah. 2021. Social media for social good? A thematic, spatial and visual analysis of humanitarian action on Instagram. *Information, Communication & Society* 24, 13 (2021), 1870–1890. https://doi.org/10.1080/1369118X.2020.1748089 arXiv:https://doi.org/10.1080/1369118X.2020.1748089

[88] Ella McPherson. 2015. ICTs and Human Rights Practice: A report prepared for the UN Special Rapporteur on extrajudicial, summary and arbitrary executions. *Centre of Governance and Human Rights, University of Cambridge* (2015).

[89] Patrick Meier. 2015. *Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response*. CRC Press, Inc., USA.

[90] UNDP Mexico. 2020. Lanza Pnud Inventario Social covid-19 para identificar iniciativas que fortalezcan respuesta Cívica de la CDMX: El Pnud en México. https://www.mx.undp.org/content/mexico/es/home/presscenter/pressreleases/2020/05/lanza-pnud-inventario-social-covid-19-para-identificar-iniciativ.html

[91] Milagros Miceli and Julian Posada. 2022. The Data-Production Dispositif. arXiv. https://doi.org/10.48550/arXiv.2205.11963 arXiv:2205.11963 [cs] type: article.

[92] Milagros Miceli, Julian Posada, and Tianling Yang. 2022. Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? *Proc. ACM Hum.-Comput. Interact.* 6, Article 34 (Jan. 2022), 14 pages. https://doi.org/10.1145/3492853

[93] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. 4, CSCW2, Article 115 (Oct. 2020), 25 pages. https://doi.org/10.1145/3415186

[94] Milagros Miceli, Tianling Yang, Adriana Alvarado Garcia, Julian Posada, Sonja Mei Wang, Marc Pohl, and Alex Hanna. 2022. Documenting Data Production Processes: A Participatory Approach for Data Work. *[forthcoming]* 1, 1 (2022), 34.

[95] Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 161–172. https://doi.org/10.1145/3442188.3445880

[96] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Rosenquist. 2021. Understanding the Demographics of Twitter Users. *Proceedings of the International AAAI Conference on Web and Social Media* 5, 1 (Aug. 2021), 554–557. https://ojs.aaai.org/index.php/ICWSM/article/view/14168

[97] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33, 4 (2020), 659–684.

[98] Asmat Monaghan and Mark Lycett. 2013. Big data and humanitarian supply networks: Can Big Data give voice to the voiceless?. In *2013 IEEE Global Humanitarian Technology Conference (GHTC)*. 432–437. https://doi.org/10.1109/GHTC.2013.6713725

[99] Michael Muller, Cecilia Aragon, Shion Guha, Marina Kogan, Gina Neff, Cathrine Seidelin, Katie Shilton, and Anissa Tanweer. 2020. Interrogating Data Science. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 467–473. https://doi.org/10.1145/3406865.3418584

[100] Michael Muller, Melanie Feinberg, Timothy George, Steven J. Jackson, Bonnie E. John, Mary Beth Kery, and Samir Passi. 2019. Human-Centered Study of Data Science Work Practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299018

[101] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300356

[102] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, and Casey Dugan. 2021. *Designing Ground Truth and the Social Life of Labels*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445402

[103] Gina Neff, A. Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science. *Big Data* 5 (2017), 85 – 97.

[104] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıciman. 2019. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data* 2 (2019), 13. https://doi.org/10.3389/fdata.2019.00013

[105] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, USA.

[106] Samir Passi and Steven Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2436–2447. https://doi.org/10.1145/2998181.2998331

[107] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 136 (Nov. 2018), 28 pages. https://doi.org/10.1145/3274405

[108] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (Nov. 2021), 100336. https://doi.org/10.1016/j.patter.2021.100336

[109] Lucy Pei, Benedict Salazar Olgado, and Roderic Crooks. 2022. Narrativity, Audience, Legitimacy: Data Practices of Community Organizers. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 328, 6 pages. https://doi.org/10.1145/3491101.3519673

[110] Kathleen H. Pine and Max Liboiron. 2015. The Politics of Measurement and Action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3147–3156. https://doi.org/10.1145/2702123.2702298

[111] Columba Teresa de los Milagros Quiñonez Amezquita. 2005. Ph. D. Dissertation. http://ri.ibero.mx/handle/ibero/867

[112] Rashida Richardson, Jason Schultz, and Kate Crawford. 2019. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423

[113] Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 92 (12 2017), 27 pages. https://doi.org/10.1145/3134727

[114] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[115] Rajeev Sharma, Sunil Mithas, and Atreyi Kankanhalli. 2014. Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems* 23, 4 (2014), 433–441.

[116] Craig Silverman. 2014. *Verification handbook: an ultimate guideline on digital age sourcing for emergency coverage*. European Journalism Centre.

[117] Alana Conner Snibbe. 2006. Drowning in data. *Stanford Social Innovation Review* 4, 3 (2006), 39–45.

[118] Jennifer Stoll, W. Keith Edwards, and Elizabeth D. Mynatt. 2010. Interorganizational Coordination and Awareness in a Nonprofit Ecosystem *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/1718918.1718930

[119] Martina Tazzioli. 2022. Extract, Datafy and Disrupt: Refugees' Subjectivities between Data Abundance and Data Disregard. *Geopolitics* 27, 1 (2022), 70–88. https://doi.org/10.1080/14650045.2020.1822332 arXiv:https://doi.org/10.1080/14650045.2020.1822332

[120] Divy Thakkar, Azra Ismail, Pratyush Kumar, Alex Hanna, Nithya Sambasivan, and Neha Kumar. 2022. When is Machine Learning Data Good?: Valuing in Public Health Datafication. In *CHI Conference on Human Factors in Computing Systems*. 1–16.

[121] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing Depression from Twitter Activity *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3187–3196. https://doi.org/10.1145/2702123.2702280

[122] Zeynep Tufekci. 2014. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ArXiv* abs/1403.7400 (2014).

[123] Janet Vertesi and Paul Dourish. 2011. The Value of Data: Considering the Context of Production in Data Economies. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 533–542. https://doi.org/10.1145/1958824.1958906

[124] Amy Voida. 2014. *A Case for Philanthropic Informatics*. Springer International Publishing, Cham, 3–13. https://doi.org/10.1007/978-3-319-05963-1_1

[125] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 915–924. https://doi.org/10.1145/1978942.1979078

[126] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2012. Bridging between Organizations and the Public: Volunteer Coordinators' Uneasy Relationship with Social Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1967–1976. https://doi.org/10.1145/2207676.2208341

[127] Amy Voida, Ellie Harmon, Willa Weller, Aubrey Thornsbury, Ariana Casale, Samuel Vance, Forrest Adams, Zach Hoffman, Alex Schmidt, Kevin Grimley, Luke Cox, Aubrey Neeley, and Christopher Goodyear. 2017. Competing Currencies: Designing for Politics in Units of Measurement. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 847–860. https://doi.org/10.1145/2998181.2998209

[128] Xiaofeng Wang, Donald E Brown, and Matthew S Gerber. 2012. Spatio-temporal modeling of criminal incidents using geographic, demographic, and Twitter-derived information. In *2012 IEEE International Conference on Intelligence and Security Informatics*. IEEE, 36–41.

[129] Matthew L Williams, Adam Edwards, William Housley, Peter Burnap, Omer Rana, Nick Avis, Jeffrey Morgan, and Luke Sloan. 2013. Policing cyber-neighbourhoods: tension monitoring and social media networks. *Policing and society* 23, 4 (2013), 461–481.

[130] Donghyeon Won, Zachary C. Steinert-Threlkeld, and Jungseock Joo. 2017. Protest Activity Detection and Perceived Violence Estimation from Social Media Images. *CoRR* abs/1709.06204 (2017). arXiv:1709.06204 http://arxiv.org/abs/1709.06204

[131] Marisol Wong-Villacres, Cristina M. Velasquez, and Neha Kumar. 2017. Social Media for Earthquake Response: Unpacking Its Limitations with Care. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 112 (Dec. 2017), 22 pages. https://doi.org/10.1145/3134747

[132] Amy X. Zhang, Michael Muller, and Dakuo Wang. 2020. How Do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 022 (May 2020), 23 pages. https://doi.org/10.1145/3392826

[133] Matthew Zook, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy* 2, 2 (2010), 7–33.