DESIGNING HUMAN-CENTERED TECHNOLOGIES TO MOBILIZE SOCIAL MEDIA DATA INTO INSTITUTIONAL CONTEXTS

A Dissertation
Presented to
The Academic Faculty

By

Adriana Alvarado Garcia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing

Georgia Institute of Technology

August 2022

DESIGNING HUMAN-CENTERED TECHNOLOGIES TO MOBILIZE SOCIAL MEDIA DATA INTO INSTITUTIONAL CONTEXTS

Thesis committee:

Dr. Christopher A. Le Dantec
School of Interactive Computing
Georgia Institute of Technology

Dr. Amy S. Bruckman
School of Interactive Computing
Georgia Institute of Technology

Dr. Munmun De Choudhury
School of Interactive Computing
Georgia Institute of Technology

Dr. Yanni Alexander Loukissas
School of Literature, Media, and Communication
Georgia Institute of Technology

Dr. R. Stuart Geiger
Department of Communication and the Halıcıoğlu Data Science Institute
University of California, San Diego

Dr. Miguel Luengo-Oroz
United Nations Global Pulse
United Nations

Date approved: May, 23 2022

ACKNOWLEDGMENTS

Completing my Ph.D. has been one of the most challenging and rewarding experiences, and it has been only possible thanks to the support, care, and friendship of an extraordinary group of people.

To my husband, friend, colleague, and mentor Benjamin, who saw in me a strength that I ignored and who has unconditionally supported me for more than a decade. Every day you remind me how lucky I am to have crossed paths with you.

To my advisor, Christopher, thank you for taking me on as an advisee, giving me the space to pursue the research I wanted, and supporting me with patience and care. I am thankful for having such supportive committee members: Amy S. Bruckman, Munmun De Choudhury, Yanni Alexander Loukissas, R. Stuart Geiger, and Miguel Luengo-Oroz. Amy, thank you for the many thought-provoking questions. Thank you, Munmun, for your feedback throughout my qualifying exams, proposal, and dissertation. Yanni, thank you for introducing me to the field of critical data studies, for always being available to talk, and for your encouragement. I would always cherish our conversations on data. Stuart, our conversations helped me clarify my ideas and better understand what I wanted as a result of the dissertation. Thank you for making me feel comfortable and valued when discussing my research. Miguel, thank you for your thoughtful feedback. I learned so much from each of our conversations, your experience working on the ground was critical for my research, and I am grateful that you shared your experiences with me.

I am thankful for the opportunity of working with the United Nations Development Programme (UNDP) Accelerator Lab in Mexico; the experience profoundly changed me as a person and transformed the course of my research. Jorge Munguía, Gabriela Ríos, Luis Cervantes, Alicia López, and Vero Quinza, thank you for receiving me to the lab and introducing me to the world of the United Nations. You all made me feel welcome and appreciated. I am glad to have witnessed how the lab started and evolved. I sincerely very much miss our Monday check-in meetings, as well as learning about all the extraordinary work you do every day. Especially, I thank my mentors during my time in the lab, Jorge and Gaby. Jorge, you were a constant source of hope and validation. Our conversations heavily influenced my research and perspectives on the intersection of technology and communities. I learned so much from you. Gaby, I am thankful I had the chance to collaborate with you. Your expertise and knowledge of conducting community research and design have been incredibly formative for me. Your creativity and care always inspire me, and I hope one day to be as good as you are telling the stories of my work.

During these years, I have been very lucky to have met extraordinary researchers that taught me and supported me along the way. To Lynn Dombrowski, thank you for being an exceptional mentor, teaching me about language justice, and for your continuous support and friendship. I always find myself remembering and putting into practice your advice. To my friend, Marisol, thank you for introducing me to new literature, helping me feel comfortable with my identity while living in the U.S, for your continuous feedback, and for supporting me in making sense of my data. You have been an incredible friend, collaborator, and mentor. I am excited to see what we will do together in the future.

I am deeply indebted to my lifelong friends and chosen family Esmeralda, Daniel, and Katy. Esmeralda, being your friend has taught me the importance of nurturing friendships and imagining other possibilities and ways of doing. You are always a source of joy. Daniel, thank you for being a friend through so many life transitions. Your friendship, knowledge, and sensibility about literature, cinema, and culture, in general, have become a true refuge. Katy, thank you for all the ways you have accompanied me through the years. Your creativity always amazes me.

Lastly, it is essential to recognize that this research could not have been carried out without the support of civil and non-profit organizations in Mexico — Los Otros Desaparecidos, UNDP - Accelerator Lab, Social TIC, CIEP — who were willing to participate in my research. Thank you very much for all your support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Social media platforms have become an established and alternative mechanism for communities to mobilize and exchange information in response to humanitarian or local crises. Due to the richness of experiences accumulated on social media platforms, this content can be valuable for civil and non-profit organizations working to address social and development challenges. At the core of my dissertation is to examine what entails not only analyzing social media data but also the implications of integrating the insights obtained from that analysis into the context of actors and institutions who might act upon those insights, such as civil and non-profit organizations.

Using social media data as evidence by institutions to inform their work entails three main challenges —accuracy, representation, and context— due to the nature of social media data. Additionally, using this type of content to inform the design of interventions and technologies that will support the studied communities entails reflecting on how we make sense of data. Within the CSCW and HCI community, there has been a growing focus on using a computational approach to establish metrics and develop tools to analyze and make inferences from social media data. However, by constraining the examination of this type of data through the exclusive use of computational techniques, there is a high risk of neglecting the social, cultural, and temporal context of the data.

In response, my fieldwork consisted in following a mixed-methods approach to understanding the underlying situations that force communities to use social media platforms as a means of organization and the implications for non-profit organizations to make social media data actionable to inform their work. Based on the findings of my fieldwork, I designed, deployed, and evaluated a toolkit addressed to practitioners working in civil and non-profit organizations interested in using data from Twitter to identify local capacities, monitor community crises, and develop interventions. The toolkit comprises computational tools that allow searching, collecting, and analyzing data from Twitter. Additionally, the

toolkit includes a manual and worksheets that guide practitioners to critically approach social media data and recognize the possibilities and limitations of this type of data by considering the challenges previously mentioned —accuracy, representation, and context.

In summary, the outcome of my research provides empirical evidence and situated tools for approaching social media data not as an independent identity but rather in light of the interplay between the online and offline behavior of the communities that produce such data. This dissertation offers two contributions to the growing body of work in HCI and CSCW invested in reflecting on the transformative character of data. First, it illuminates the large ecosystem of norms and practices of multiple actors, infrastructures, and databases that we need to consider to mobilize data from online platforms into institutional contexts. Second, the design of the toolkit proposes an actionable example of how to promote a situated examination of data. While my research has focused only on examining data from social media platforms, the contributions of my work are meaningful in the broader context of data-centric technologies. As we continue to deploy this technology, it is imperative to interrogate the assumptions and biases encapsulated within those technologies, specifically the data that feed them, and how they impact our understanding of human networks and communities.

CHAPTER 1
INTRODUCTION

Data have become the primary fuel for contemporary technology, feeding the algorithms that make possible AI systems, offering predictions about human behavior at scale [1, 2], algorithmically shaping user experience, economic, and labor decisions [3, 4, 5], and charting paths of individual and collective action [6, 7]. The focus of my research is on a specific kind of data: user-generated content published on social media platforms[1].

In the last decade, the fields of Social Computing and Computational Social Science have increasingly used societies' digital traces to track people's activities, identify patterns and draw interpretations from them [8, 9]. These disciplines use computational techniques to establish metrics and develop tools to analyze and make inferences from social media data, under the assumption that these techniques are an objective and unobtrusive approach to examining human behavior at a scale. Some examples include, characterizing online communities [10, 11], predicting mental health states and deviant behavior [12, 13], and assessing rates and prevalence of criminal activities [14, 15]. However, by constraining the examination of this type of data through the exclusive use of computational techniques and abstract approaches there is a high risk of neglecting the social, cultural, and temporal context of the data [16].

Data, communities, and places are always intertwined. If the ultimate objective of analyzing social media data is to inform the design of interventions and technologies that will support the studied communities, we need to develop mechanisms to conduct a more situated analysis of social media data. To conduct a situated approach is imperative to anchor the analysis of user-generated content on the communities that produce them beyond

[1]From here on, I refer to the data produced on social media platforms interchangeably as social media data, digital traces, and user-generated content.

the online context.

In my dissertation, I am concerned with examining what it entails not only to analyze social media data but also the implications of operationalizing the integration of the insights obtained from analysis into the context of actors and institutions who might act upon those insights, specifically civil and non-profit organizations. The following account describes my path in developing a set of tools that support the mobilization of data from their platform of production into the institutional context of organizations. In developing these tools, I examine the production of social media data in light of the online and offline communities they are part of. To this end, I employed methodologies to capture fundamental aspects of the social context in which communities are turning to online spaces to organize and exchange information. Then, I partnered with a civil organization to examine the implications of making social media data actionable in a determined context.

To design these tools, I have draw from the field of Critical Data Studies, which interrogates the social and political process and assumptions involved in data creation, curation, and unequal power dynamics in the pipeline of imposition of meaning and classification of data [17, 16, 18, 19, 20]. This increased volume of literature attests that, first, data in itself has become an object of study [21], and second, underscores the need to understand data as situated and socially constructed, rather than an independent entity [19].

My research is in conversation with emerging perspectives in the fields of Computer-Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) that have begun to reflect on data's transformative character. Examples of these reflections include critical analysis on the creation, documentation, and exploitation of datasets [22], data practices and power dynamics on data labeling [23], and the various human interventions involved in data science practices [24, 25, 26], which in many cases begin even before the data reach the computer [27]. These examples are beginning to uncover the social and technical actors in the chain of data production and use.

Table 1.1: Summary of studies and research questions

Study	Methods	Research Questions
Use of social media platforms to collect and exchange information of crime in high violence context. **S1**	Interviews, online ethnography	**RQ 1:** How have people adopted Facebook groups to collect and exchange information about crimes in high violence contexts? **RQ 1a:** What practices do people adopt to establish trust and participation in online communities where citizens exchange information about crimes? **RQ 1b:** How do the administrators of these online communities scaffold the information collected? **RQ 1c:** What are the challenges of the administrators of these online communities to scaffold the collected information?
Identifying signals in social media data streams to leverage as a resource for NGOs developing evidence to support their work. **S2**	Mixed methods approach: Online ethnography, content analysis, computational methods	**RQ 2:** How can meaningful information be identified on social media platforms for use as evidence by civil and non-profit organizations seeking legal and community interventions? **RQ 2a:** How to identify evidence in social media data that can potentially be linked to official data records? **RQ 2b:** What attributes of social media data qualify them as evidence? **RQ 2c:** How can we design tools that address the volume challenge of finding evidence across social media data to inform the work of NGOs?
Examining the methodological challenge of making social media data legible for a civil organization. **S3**	Mixed methods approach: Online ethnography, content analysis and computational methods	**RQ 3:** What are the practices involved in mobilizing social media data from their site of production to the institutional context of non-profit organizations? **RQ 3a:** How do NGOs practitioners define what social media data counts as evidence for their work? **RQ 3b:** What are the challenges of integrating insights from social media data into NGOs' work?
Design, implement, and evaluate the toolkit *Bitácora*. Provide a set of guidelines based on the evaluation of the toolkit. **S4**	Mixed methods: Prototyping, interviews, and workshops.	**RQ 4:** What type of tools are needed to support practitioners in mobilizing social media data from online communities into their institutional context? **RQ 4a:** How do we design tools that promote situated and critical perspectives when organizations practitioners integrate social media data into their work? **RQ 4b:** What are the necessary guidelines to support organizations' practitioners in evaluating whether social media data is appropriate to inform their work? **RQ 4c:** When mobilizing social media data to NGOs, how do we design supportive tools that document the context of the data's production?

1.1 Research Questions

To address my overarching research interest in examining how to mobilize social media data into institutional contexts, this dissertation addresses the following research questions. Table 1.1 summarizes these phases, their studies, and the research questions they respond to in connection to the larger issues this dissertation explores.

1. **RQ 1:** How have people adopted Facebook groups to collect and exchange information about crimes in high violence contexts?
   (a) **RQ 1a:** What practices do people adopt to establish trust and participation in online communities where citizens exchange information about crimes?
   (b) **RQ 1b:** How do the administrators of these online communities scaffold the information collected?
   (c) **RQ 1c:** What are the challenges of the administrators of these online communities to scaffold the collected information?
2. **RQ 2:** How can meaningful information be identified on social media platforms for use as evidence by civil and non-profit organizations seeking legal and community interventions?
   (a) **RQ 2a:** How to identify evidence in social media data that can potentially be linked to official data records?
   (b) **RQ 2b:** What attributes of social media data qualify them as evidence?
   (c) **RQ 2c:** How can we design tools that address the volume challenge of finding evidence across social media data to inform the work of NGOs?
3. **RQ 3:** What are the practices involved in mobilizing social media data from their site of production to the institutional context of non-profit organizations?

(a) **RQ 3a:** How do NGOs practitioners define what social media data counts as evidence for their work?

(b) **RQ 3b:** What are the challenges of integrating insights from social media data into NGOs' work?

4. **RQ 4:** What type of tools are needed to support practitioners in mobilizing social media data from online communities into their institutional context?

   (a) **RQ 4a:** How do we design tools that promote situated and critical perspectives when organizations practitioners integrate social media data into their work?

   (b) **RQ 4b:** What are the necessary guidelines to support organizations' practitioners in evaluating whether social media data is appropriate to inform their work?

   (c) **RQ 4c:** When mobilizing social media data to NGOs, how do we design supportive tools that document the context of the data's production?

1.2 Contributions

At the core of my doctoral research has been understanding the underlying situations that force communities to use social media platforms as a means of organization and the implications of connecting the data of those collective experiences with institutions that might leverage and act upon them. My doctoral research started by understanding the context and the communities that crowdsourced content in the context of high violence such as the *War on Drugs* and the ongoing crisis of Human-Rights violations in my home country, Mexico. With each additional study I extended my research into offline and institutional contexts to understand the larger ecosystem and the multiple actors involved in leveraging such valuable local knowledge embedded in the content of social media platforms. On doing this transition I shifted the focus of my research from technologies to address human rights violations to investigate the role of data practices of government and organizations and human interpretation into making social media data actionable.

Because data are the product of social relations, understanding the potential of social media content means examining the data practices of those who produce the content and those organizations that can use the data. By approaching social media data not as an independent identity but rather in light of the interplay between the online and offline behavior of the communities that produce such data, my research contributes to the growing body of work in HCI and CSCW invested in reflecting on the transformative character of data.

- **Empirical Insights Into Collective Online Organization:** Communities' adoption of Facebook and Twitter in Mexico to manage and make sense of data about crimes, human rights violations, and local crises is an outstanding case to understand the potential and limitations of using social media data as evidence to inform the interventions of organizations seeking to address community crises. My research provides empirical insights into the data practices that these communities engage with to collect, curate, and publish information about safety [28], confront the ongoing crisis of missing people in Mexico [29], and address the social and economic challenges of the COVID-19 health crisis. The findings of the three studies tease apart the implications of using social media data as evidence to augment and contest official records of the ongoing crisis of disappearances in Mexico, monitor communities' resilience, and inform the work of civil organizations.

  Taking together the three accounts illuminate the large ecosystem of norms and practices of multiple actors, infrastructures, and databases that we need to consider to mobilize data from online platforms into institutional contexts. Moreover, the findings of my research contest the assumptions that social media data can be examined as an independent entity.

- **Situated Human-Centered Tools:** As a result of my empirical research, I designed a toolkit that comprises computational and qualitative tools that encourage and advise practitioners from civil and non-profit organizations to reflect on the dynamics encoded in the process of collecting social media data to be later situated in institutional contexts. Specifically, the tools and guidelines I designed inform and support practitioners in the process of integrating the input of the communities behind social media data and their counterparts in offline settings. By developing tools that center community perspectives and the context of data production, my research contributes to transforming the research practices that surround the construction of data and the harboring of interpretations, to surface the possibilities and limitations of data. Lastly, I am anchoring the design of the tools and guidelines in the community's perspective to raise awareness about the values that can be derived from data and challenge the notion that data-related work is neutral.

In designing and evaluating the toolkit, I provide an example of how to design sociotechnical systems that address the challenge of retaining the context of data at scale by advocating for an understanding rooted in the communities that produce and consume the data.

Additionally, the toolkit illustrates the valuable perspectives that Critical Data Studies provide to design and deploy more equitable technologies. Drawing from this discipline, the toolkit encourages practitioners to approach user-generated content from a critical perspective by promoting documentation practices and reflexivity when analyzing and integrating data from Twitter into their work.

1.3 Overview of Dissertation

This dissertation explores the previously mentioned research questions through a mixed-methods approach and across four phases of research and design. The rest of the dissertation is organized as follows.

Chapter 2 provides the contextual background in which I have conducted my research, including an overview of the current human rights crisis in Mexico and a description of my partnership with the Accelerator Lab in Mexico City.

Chapter 3 outlines the related works and the theoretical perspectives that have informed and shaped the design and direction of my research.

Chapter 4 describes the first stage of my research, which consisted of examining how people adopt Facebook groups in Mexico to collect and exchange information about crimes in high-violence contexts. The findings of this study led me to identify that despite the value of the crowdsourced data from these online communities, the fact that the data was unstructured and difficult to verify limited the way administrators and contributors could use it to address issues of crime and public safety. These challenges surfaced the need to move from individual to systemic practices by establishing mechanisms that facilitate the process of making this data actionable.

Chapter 5 reports on the second stage of my research focused on developing mechanisms to identify meaningful information in social media data and transform it into evidence for organizations seeking legal and community interventions. Following a mix-methods approach, I examined data from Facebook groups where people reported cases of missing people. Building on the results, I established a data gap between the number of cases officially recorded in the government database of missing people in comparison to the official accounts of missing people circulating on Facebook. Additionally, I identified seven characteristics of social media data that shape the process of systematically using social media data as a robust proxy measure for under-reported events, such as human-rights violations. However, quantifying social media data and linking it to official data records are not enough to inform organizations seeking to make community interventions. To transform the knowledge of online communities into usable and actionable evidence, it is imperative to understand organizations' definitions of evidence and organizational practices to render social media content actionable.

Chapter 6 describes the third stage of my research, focused on examining what entails making social media data legible for NGOs. Following a mix-methods approach and partnering with the team of the Accelerator Lab in Mexico, I examined data from Twitter to characterize how citizens, local governments, and grassroots organizations in Mexico City collaborated to address the social and economic impact of the COVID-19. The process of making this characterization entailed multiple cycles of collaborative interpretation with my partners to categorize the data from Twitter. The findings of this study highlight that using computational tools for extracting and making sense of social media data are not purely technical activities. Instead, these activities involve countless trade-offs, decisions, and assumptions when categorizing and interpreting the data, concealing a set of values and interests of those involved in doing the work of data. To design computational tools that support mobilizing social media data into an institutional context demands acknowledging what counts as evidence for the institution while considering the attributes of social media content. More importantly, aiming for an organization or institution of any type to integrate social media content as an alternative source of information implies assimilating that content into the data ecosystem in which the organization already operates.

In consideration of these needs, in the fourth stage of my research, reported in Chapter 7, I developed a toolkit, name *Bitácora*. The toolkit aims to support practitioners of non-profit and non-governmental organizations to integrate social media content into their work while promoting situated and critical perspectives on the implications of using such content. The toolkit comprises a set of guidelines and recommendations to help practitioners evaluate whether social media data is appropriate for them to use and analyze them from a qualitative perspective.

Chapter 8 reports on the design of evaluation and findings of deploying the toolkit with the Accelerator Lab and with two non-profit organizations in Mexico.

Lastly, Chapter 9 concludes my dissertation with a reflection on the implications of my research and potential future directions.

CHAPTER 2
BACKGROUND & PARTNERSHIPS

2.1 Research Site

In this chapter, I provide background information on the conditions of violence that have affected Mexico, the site of this research, during the last decade. Then, I situate this dissertation in light of the growing body of computational research that operationalizes social media data to characterize online communities and predict health conditions, crime, and human behavior. Lastly, I conclude with a revision of existing methods and frameworks from social science and humanities concerned with examining the processes that make data transportable and legible after data are moved and disconnected from their site and discipline of production [30, 16].

2.1.1 Data in Mexico

I conducted the research of this dissertation in my home country Mexico. Due to the War on Drugs that started in 2006, Mexico has been experiencing a renewed crisis of human-rights violations and extreme violence coordinated by organized crime and state security forces [31, 32, 33, 34, 35]. The ongoing crisis has given rise to two data-related converging factors –missing data and lack of data standardization– making Mexico an appropriate place to examine how communities engage with technology as a way to organize and address local crises.

Missing Data

Due to decades of impunity and Mexico's long and complicated history of police and political corruption, much of the Mexican population distrust the justice sector. The resulting lack of trust creates an environment that discourages citizens from filing criminal complaints but also prevents authorities and non-government organizations from having an accurate understanding of the magnitude of the situation and from developing appropriate policy responses [36, 37]. In some cases, victims fear reprisals against their families or themselves and do not inform authorities of crimes. Citizens are also unlikely to report incidents because of a perceived hostility from the authorities and because reporting procedures are onerous and ineffective [37]. This lack of reporting contributes to what many Mexican citizens and institutions refer as the *dark figure,*[1] which is the percentage of crimes not reported or reported crimes that did not result in an investigation [31, 37]. According to the National Survey on Victimization and Perception on Public Safety, [2] the dark figure in 2020 was 93.3% nationwide [38]. Results of this survey, also show that the main reasons preventing victims of crime from reporting are rooted in their relationship to the authorities. In 2020, 60.7% of victims reported not filing complaints because of reasons related to the authorities. Of these, 33.9% of victims considered filing complaints a waste of time and 14.2% stated not trusting the authorities as their rationale for not reporting a crime [38]. In face of these challenges, scholars speculate that law enforcement is low due to governmental corruption and a lack of sufficient data collection and evidence for crimes, resources to pursue litigation, or effective policies and deterrents [39].

Lack of data standardization

In addition to the missing data, there is a lack of standardization in how each state and municipality in Mexico classify crime data. As a result, prosecutors are often forced to improperly classify crimes based on the categories that exist for their state and different organizations use different sources of evidence to calculate homicide statistics [35, 31]. While the National Institute of Statistic and Geography (INEGI) takes into account both intentional and accidental deaths, and bases its findings on administrative records of homi-

[1]In Spanish: La cifra negra
[2]In Spanish: Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE).

cides from Civil Registry Offices supplemented with data from other authorities [40], the National Public Safety System solely relies on preliminary investigations or investigation files initiated in civil courts. In the latter, each investigation is treated as one data point; yet, since a single homicide investigation can be linked to multiple victims, such data often understates the actual number of victims. Furthermore, disparities on how municipalities report and record crime data exist too. While some maintain data electronically and online, others rely on paper copies that are inaccessible to the public.

Because there is no single standard across Mexico, prosecutors are often forced to improperly classify crimes based on the categories that exist for their region. In some jurisdictions, 'enforced disappearance' is often reclassified as 'kidnapping' because the former is not a recognized category [31]. Indeed, of the 32 Mexican states, only eight have included enforced disappearance in their criminal code; of those that have, each has defined it differently [34, 41]. This blurring of categories effectively merges or obscures differences [42]. The resulting data are thus muddied, with some crimes being over-reported while others are under-reported.

Violence against the press

In addition to the discrepancies and gaps in data about crime and safety, human rights organizations have consistently reported an increased hostility toward the free press [43]. While censorship and a general lack of freedom of expression are not entirely new in the country, since the beginning of the War on Drugs, censorship, and violence against journalists have increased considerably [44].

In Mexico, the impunity for any category of homicide exceeds 90%. In the case of journalists, the organization Article 19 [3] calculates that there is a 99.13% rate of impunity and unsolved crimes since many of these cases are intentionally delayed in prosecution for years. For example, in 2017 alone, there were 14 journalists killed in Mexico, from

[3] Article 19 is an independent and non-partisan organization that promotes and defends the rights of freedom of expression and access to information for all people. Website: https://articulo19.org

which at least six murders were confirmed as direct reprisals for their work. None of these cases led to a conviction [45]. In 2017, Mexico was the deadliest country for journalists outside conflict zones such as Iraq and Syria [46]. Then, at the end of 2020, the Annual Review of the Reporters Without Borders revealed that Mexico continued to be the most dangerous country in the world for journalists among countries being considered *to be at peace*, with at least eight cases of journalists executed, for investigating links between organized crime and politicians [47]. From 2000 to date, Article 19 has documented 131 murders of journalists in Mexico potentially related to their work. Of which 47 cases occurred during the term of ex-president Enrique Peña Nieto (2012-2018), and 11 happened since Andrés Manuel López Obrador began his term in December 2018 [43].

In consequence, not only are crimes not being reported to the authorities, but they are also being kept out of the public discourse through violence against journalist. The increase violence in the country have forced citizens to look outside established institutions in order to act for their own public safety and to begin to address the systemic challenges faced in both urban and rural parts of Mexico. These conditions help explain the rise of social media platforms as sites that enable communities to organize, report and publicly display stories and experiences about crimes that would otherwise go unreported.

2.2 Partnerships

My research has been possible through collaboration with organizations in Mexico. While I was conducting the second research study [29], I partnered with the Non-Governmental Organization (NGO) *Los Otros Desaparecidos de Iguala* [4] founded in 2014 in the state of Guerrero in Mexico by relatives of missing people. This organization is one of many collectives that coordinates search squads and recover bodies in clandestine graves. In the last decade, grassroots organizations like this have sprung up all over the country.

At the time I started collaborating with *Los Otros Desaparecidos de Iguala* the focus

[4] In English: The Other Missing of Iguala

of my research was still on understanding how to effectively mobilize social media data to confront the human rights crisis in Mexico. A traditional approach to addressing structural social ills like crime and violence is to understand the magnitude of the problem through quantification to inform policy and legal responses that intervene at the appropriate scale – local, state, nation. My initial expectation working with this organization was using social media data to facilitate the quantification of missing people. However, for those organizations searching for missing people, the data has become secondary. Instead, they care about working on the ground, searching for clandestine graves to identify as many bodies as possible. NGOs' response is understandable if we consider that the disappearance of people in Mexico has been an uninterrupted practice since the 1960s [48, 49, 50].

2.2.1 Collaboration with the Accelerator Lab in Mexico

In January 2019, I joined the first cohort of the United Nations Global Pulse Data Fellows initiative. As part of this fellowship, I was assigned to work with the Accelerator Lab Network. Then, in January 2020, I was assigned to work with the Accelerator Lab in Mexico City and collaborated with them until September 2020 as part of my role as Data Fellow. Then, I was hired as an intern in April 2021 for a first term of six months and then for a second term of three months. Overall, I collaborated with the lab for 18 months.

The Accelerator Lab Network of the United Nations Development Programme (UNDP) consists of 91 labs that support 115 countries. Unlike previous approaches to address development problems, the Accelerator Lab focuses on working closely with local stakeholders to identify community-level solutions that have the potential to accelerate development. Among the priorities of the Accelerator Lab is establishing new methods to identify innovative sources of information that facilitate high-level decision-making on complex development problems. Accelerator Labs work in an iterative cycle consisting of sensing, exploring, testing, and growing, and each lab consists of three members with the following titles: head of solutions mapping, head of the exploration, and head of experimentation.

As a data fellow and intern, I collaborated with a broad range of projects focusing on two main challenges: 1) strengthen communities' capacity to recover from crisis through social capital, and 2) stop gender violence through new approaches. I spent most of my time collaborating with the team members in charge of the solutions mapping and the exploration. My day-to-day work ranged from note-taking and preparing materials for meetings organized by the Lab with various stakeholders, conducting literature reviews, writing and synthesizing reports and blogs for publication, and in the last months helping with the design evaluation of technology [51, 52].

The experience of working with the Accelerator Lab allowed me to see how the organization builds and maintains partnerships with multiple stakeholders, including government agencies, the private sector, academia, and other units with the UNDP. Due to the length of our collaboration, I witnessed the beginning and conclusion of various projects, the strengthening of their collaborations, and the impact of their interventions.

Working with the Accelerator Lab profoundly transformed the course of my research. During spring 2020, when the COVID pandemic started, the lab launched the exploration *COVID-19 Social Inventory* to map citizen initiatives that respond to the challenges imposed by the pandemic [53]. As a Data fellow, they allowed me to conduct a study in parallel that consisted in analyzing Twitter data as an alternative source of narratives that could reflect unusual citizen strategies to address the COVID-19 crisis[5]. Conducting this study in collaboration with them compelled me to rethink what it entails to mobilize social media data from online into an institutional context. The long-term partnership helped me to recognize the role of data and organizational practices in making this integration of social media content into their work.

[5]This investigation corresponds to the third study that I conducted and that I describe in chapter 6

CHAPTER 3

RELATED WORK

3.1 Examining the Networked Public Sphere

Social media platforms have proven to be an effective place for developing new ways of organizing through the accumulation of crowdsourced information [54]. Due to their unique affordances, like providing insights about the local narratives, capabilities, and expertise from citizens, social media platform data have the potential to become a tool that works at scale to gather evidence useful for addressing human rights and local crises [55, 56, 57]. In the last decade, there has been an increasing body of work focused on examining how people turn to social media platforms during an acute crisis [58, 59], while searching for emotional support [60, 61], or to address local issues [62, 63, 57]. However, effectively operationalizing the integration of social media data from online to institutional contexts to be used as evidence to inform community and policy interventions remains an understudied area within the discipline of HCI.

In this section, I provide an overview of Crisis Informatics and Digital Civics, research fields concerned with the role of social media platforms as a way of a collective organization either to inform the large-scale emergency response to natural disasters or to enable social movements. Next, I describe an emerging area of Social Computing that focuses on examining user-generated content to characterize online communities, understand mental health, and forecast human behavior. I conclude this section by outlining some of the main limitations of using social media data to characterize communities and human behavior raised by scholars from Science and Technology Studies (STS).

3.1.1 Crisis Informatics and Digital Civics

The emergence of social media platforms and crowdsourced data create a new, distributed, and accessible set of tools for the public to organize collectively in the face of natural disasters. This turn has given rise to the field of Crisis Informatics, which focuses on examining how communities use social media platforms and data to cope with crises [64, 65]. The majority of crises that have been of concern for Crisis Informatics are natural disasters [66, 67, 68], human-made crises such as wars and terrorist attacks [69, 70, 64, 71] and, chemical and viral disease crises [72]. The large body of work on Crisis Informatics indicates that in the aftermath of natural disasters, social media platforms are no longer just emerging tools, but they have become a routine part of crisis response and a formalized infrastructure for various tasks such as collecting witness accounts, monitoring the media, and providing updates during crisis responses [65, 58, 73, 74].

Previous research has shown the different approaches of aligning social media with other socio-technical systems to perform information work after disasters. Daily and Starbird describe how individuals stitch platforms together to support information needs, and due to the flexibility of the platforms, as information needs evolved the actors involved reconfigured the organization of the platforms to meet their needs. Additional key practices include amplifying information, helping with damage assessment, and monitoring public safety [58].

One of the difficulties in relying on social media to collect and share information during a crisis is the bridge between those participating via social media and those in institutional roles that need to coordinate across multiple service boundaries. On the one hand, more people can provide aid indirectly by organizing funding and supplies and information [66, 54]; on the other hand, the fact that these activities take place on social media limits and shapes the available data both in terms of who produces it and who has access to it.

Social media platforms have also enabled new forms of organizing social movements, in which the relations between people are instantiated and mediated by an array of data, computing interfaces, and local contexts. These new forms of organization outline the emerging field of Digital Civics, which examines the role of technology in supporting citizen engagement and community organization. In recent years, Digital Civics has focused on exploring the use of technology when supporting community organizations. These interventions have used a variety of digital technologies such as social media [62, 75, 76], mobile and web platforms, civic games [77, 76, 78, 79], and hackathons [80, 81, 82]. The goals for these interventions varies from supporting broad participation in local government and institutions, to constructing social movements that empower communities in shaping their civic life (e.g., [83, 84, 85, 62, 86, 87, 76, 88]). Additionally, researchers have examined ways to motivate different types of users to gather data about their neighborhoods through interfaces that focus on increasing data gathering [89], the offering of rewards [90], or creatively deploying game mechanics [91].

While each of these examples illustrate particular strengths and encourage different modes of supporting local communities, they also present challenges related to the kind of data that communities and NGOs might mobilize. For example, in the case of blogging and networking sites that support collaboration [76], communities both generate the data and interpret them. In other cases, communication is mediated by the partnership between different stakeholders towards shaping a community and guiding the provision of public services, brokering trust as well as data. These interactions matter because they shape how NGOs and communities take action using different kinds of data, and on how information signals and organizational capacities are amplified or dampened by those interactions [54]. Across these diverse projects, it is possible to outline the ways in which forms of data production intermingle to articulate a shared set of commitments between a given community and the organizations working to support specific policy or social outcomes [92]. These commitments in turn create new capacities that enable communities to take action [93], and contribute to community resilience [85].

3.1.2 Online Digital Traces to Characterize and Predict Human Behavior

In recent years, using social media data as a reliable source of people's values, thoughts, and emotions to describe, quantify, and predict future human behavior has increased [94, 95, 96]. This trend is motivated by the assumption that social media is a rich repository of data that enables a naturalistic, unobtrusive, and objective approach to examine collective behavior [97, 98, 99]. Thus, the primary focus of this research area consists of characterizing online communities [10, 11, 13] and quantifying and predicting human behavior based on the analysis and interpretation of social media data [100, 101, 102].

Characterizing Online Communities to Predicting Human Behavior

Using textual and social network information has been extensively used to track human interactions and infer behaviors at scale, including dietary habits, human values, aspects of ideology, and mental health states and psychological well-being such as deviant behavior, stress expression, and psychological impact [10, 11, 13, 103, 104, 96, 105, 106, 107, 108].

Characterizing online communities has proven to be an alternative to provide a rich understanding of the misappropriation of social media platforms for unhealthy purposes. An example is the work of Pater conducting a content analysis of online eating disorder (ED) user-generated content and their associated hashtags and media across platforms including Tumblr, Instagram, and Twitter. Through this characterization of activities, the authors examined the role of social media platforms in supporting networks that promote negative health purposes. Their results included a comprehensive corpus of ED-terminology, as well as evidence of members of the network using social engineering techniques in response to the platform's censorship [10].

Furthermore, to evaluate and inform the design of appropriate content moderation strategies, previous research examined user-generated content that was public and later removed. Research findings showed that removed content reflected more dangerous actions such as self-harm tendencies than content that remained public. Additionally, the results provided a set of characteristics in captions and taggings of posts on Instagram that predict whether a post will be removed [12]. Similarly, to evaluate the effectiveness of content moderation strategies of pro-ED communities on Instagram, researchers characterized deviant behavior by conducting large-scale quantitative studies. Results of this research indicated that the pro-ED community adopted nonstandard lexical variations of moderated tags to circumvent Instagram's moderation policies, showing that Instagram's strategy to remove pro-ED content was ineffective at preventing the dissemination of pro-ED behavior on the platform [13].

Using social media content to examine mental health states among specific populations has been growing in the last decade, partly because the stigma associated with disclosing mental illness prevents researchers from collecting data, making user-generated content an unobtrusive alternative data source. On the other hand, using computational techniques based on linguistic analysis and psychology to analyze social media data has reinforced the notion that this type of data enables reliable monitoring of mental health states while minimizing bias [103, 104]. Health behavior inference from social media has been conducted in a variety of ways, including quantifying levels of mental illness severity, predicting the likelihood of individuals to engage in mental health discussions, and forecasting behavioral changes such as the risk of postpartum depression in new mothers.

In most cases, the characterization of online communities aims to inform the design of models, methodologies, and tools capable of monitoring behavior at unprecedented scale and deriving specific characteristics of communities to predict real-world behavior. For instance, De Choudhury et al. built predictive models using Twitter posts to understand and predict upcoming behavior and mood of new mothers. To this end, the authors used four types of measures to characterize the behavior and mood of the new mothers, including engagement, ego-network, emotion, and linguistic style. Their results show a predictive model capable of classifying mothers that will change their behavior with an accuracy of 71% when using only prenatal observations, once adding additional data from periods of postnatal data, the predictive accuracy rises to 80% [100].

In other cases, new methodologies have been developed to help discover vulnerable groups faster in a data-driven manner. One example is the first to quantify and forecast the levels of mental illness severity (MIS) in social media [11]. In this study, the authors provide a new method to predict future MIS in users who share pro-ED content on Instagram. By examining a dataset of 26M posts from 100K users on Instagram who post content related to the pro-eating disorder, authors found that even though most of the posts bearing low levels of MIS, those who share pro-ED content on Instagram show a trend of increasing MIS in their content over time. Their results show that since 2012 there has been an increase of users whose content expresses high MIS. Additionally, the authors found that they were able to forecast user's manifested MIS over up to eight months into the future based on the MIS previously observed in their past content. Similarly, researchers have leveraged social media data to identify factors related to the likelihood of recovery of eating disorders such as anorexia. For example, Chancellor et al. examine the effectiveness of social media platforms in promoting recovery from anorexia [109]. Using a dataset of over 68M posts and 10K users that self-identify with anorexia, the authors developed a statistical model based on survival analysis to predict the likelihood of recovery over time of those users who suffer anorexia.

Crime Systems

A growing body of computational research has begun to focus using social media data to assess rates and prevalence of criminal activities and violence based on social media data. Examples of this work include: using protest images as a source to characterize and estimate violence during public demonstrations [110], leveraging Twitter data to better understand patterns in incidents of crime and creating classification models to automatically label incoming social media data [111]; using Twitter data drawn from local news agencies to predict hit-and-run vehicular accidents and breaking-and-entering crimes [14]; and applying linguistic analysis and statistical topic modeling to spatiotemporally tagged Twitter to improve crime prediction performance versus a standard approaches based on kernel density estimation [112, 113]. Additional work has used Twitter data to augment and assess vehicular descriptions linked to crime [114]; forecasting likely changes in crime over time using a number of different social and environmental factors [115, 15, 116]; and estimating the population at risk of crime (such as street robbery) in a given place and time by using geotagged traces of Twitter postings [117].In this realm of research, social media data are used to augment structured data analytics and prediction models. Previous research built out a variety of techniques that use spatio temporally tagged data from social media to create crime prediction models [113, 14] and to understand patterns of occurrence [111, 116, 115].

Ultimately, what this body of research intends to demonstrate the potential of social media as a data source for assessing and predicting crime. However, this research has been conducted without considering both the offline context of the social media data and the particularities and limitations of "ground truth" which rely on existing robust data sources, established metrics, and rich analytic tools. The challenges of how this class of unstructured, crowdsourced, and voluntary data might be deployed to address gaps in official data sources, rather than augmenting those sources, remains an understudied area of empirical research.

3.2 Theoretical Perspectives and Approaches

The research in social computing that has focused on establishing metrics and developing analytic tools to analyze and make inferences from social media data assumes using computational methods as an objective and unobtrusive approach to examining human behavior at a scale. This approach disregards the role of human interpretation in the data analysis and neglects to connect the insights obtained from social media with stakeholders that could leverage that knowledge. At the core of my dissertation is addressing these omissions by examining the implications of the role of human interpretation and context when leveraging the content from social media platforms to inform the interventions of NGOs and civil society organizations. To inform this inquiry, I draw from the field of critical data studies, which argues for understanding data as situated and socially constructed rather than as an independent entity. 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.

3.2.1 Implications of data re-use

Using social media data to inform the design of interventions, methodologies, and tools entails making the content legible to multiple stakeholders, which demands removing such content from its platform of origin into a different context. Scholars from STS and Critical Data Studies have examined the challenges of making data legible when they are disconnected from their producers, the obstacles of data sharing and reuse across scientific disciplines, and how these challenges are bound to the context of data production [30, 118]. For instance, in the context of e-science, the authors examined the data sharing practices within natural and social science groups to facilitate the reuse of data through four case studies that varied on their discipline, data collection, and use of technology [30]. As these studies demonstrated, in the process of making data legible to other stakeholders or contexts, certain types of data can be neglected [22]. Carlson et al. found that quantitative disciplines are better equipped than anthropology when formalizing their data into a configuration that enables reuse. For anthropologists, data reuse would entail an epistemological challenge, because their primary goal is to provide thick descriptions of particular contexts, whereas the formatting requirements for data reuse demand condensing data to make it more generalizable. By contrast, the quantitative disciplines find it more accessible to transform their objects of study into transportable and intelligible data forms, such as visualizations and simulations [30].

The concept of data reuse assumes that knowledge can be disembedded from its place of production and carried into a different context. However, as Carlson et al. showed, this assumption is neither supported by the practices of e-science nor by other disciplines without involving an epistemological shift [30]. To partially address the challenge of data reuse when analyzing social media content to inform the design of interventions and tools, it is imperative to examine the data collected from social media platforms in their unique political, cultural, and historical context of production and circulation. Otherwise, the documented experiences in those platforms lose meaning and value [16].

Recognizing Context of and Representation within Data

Understanding and maintaining the context of social media data is a complex task that requires effort and a kind of care [63]. The moment we extract social media data from their place of production, we strip them of their conversational context [44]. Human intervention is then necessary to appropriately contextualize the meaning of the data. Moreover, examining data from social media platforms as a social construct entails recognizing the interplay between online and offline people's interactions. Since the data from these platforms reflect solely online interactions, the conclusions we can make from such content are restricted.

An additional aspect to consider when analyzing social media data is the completeness of datasets in terms of representation. Although social media data reflect a range of behaviors, ideas, and opinions, it would be a mistake to assume such content as complete datasets [16, 18]. Depending on the tools and platforms used to collect data, some communities and perspectives become more visible than others. Moreover, regardless of how large the datasets we might collect from social media platforms, these data will always be biased due to the demographic sample, limiting whose experiences, needs, and interests are visible for analysis. In consequence, these constraints of representation will always affect the inferences we can make from the data.

3.2.2 Human Intervention and Subjectivity

Previous work in the fields of CSCW and HCI has examined the role of human intervention and subjectivity associated with the work practices required to bring data into computing systems, challenging the notion of objectivity in data-related work and contesting the neutrality of these systems. Pine et al. suggest that human intervention begins even before the data reach the computer [27]. Using two case studies, the authors note that human-computer interactions with data start with the measurement interfaces, which are the artifacts and practices that give rise to the smallest units of quantification to create data, algorithms, and other forms of representation [27]. These interventions continue through the process of data collection [119, 24]. In this respect, Feinberg provides an overview of how the practices involved at the stage of data collection across different disciplines are *interpretively flexible* despite the continuous effort to implement standardized protocols and vocabularies that enforce consistency [119]. Through this work, Feinberg shows that interpretive judgment is always present in data creation; moreover, Feinberg shows how human intervention actively shapes the stages that follow data collection [119].

By providing a set of dimensioned interventions, Mueller et al. show that human mediation is constantly present across all the work practices in data science, including, discovery, capture, curation, design, and creation. These interventions influence the definition and interpretation of data, affecting the act upon and intervention in the world based on that data [24]. Furthermore, these interventions are not a neutral mechanical sequence of rules but instead are the result of situated work that demands mastering views of the worlds and tools, what Passi and Jackson refer to as *data vision* [25]. Data vision is "the ability to organize and manipulate the world with data and algorithms" which results from a professionalization process shaped by a community of practice [120, 25]. Taken together, this work provides a more nuance understanding of the craft of working with data, showing how it requires situated decisions and relies on applying and adapting prior knowledge on behalf of those working with data and algorithms to produce knowledge.

3.2.3 Methodologies for Data Migrations, Journeys, and Translations

Research on social science and humanities has developed methodologies and frameworks to examine the processes that make data transportable and legible after data are moved and disconnected from their site and discipline of production [30, 16].

One of such methodologies is the *data journey*, which consist of examining 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. Using as a case study the data that travel through the UK's weather infrastructure, authors of the methodology further examined the data journey through projects, datasets, and individuals. As the authors followed the trace of the data, they identified the socio-material factors that constitute data and capture the way social worlds are becoming interconnected and interdependent through data production, distribution, and use [121]. As data moved between different sites of practice, the *data journey* methodology surfaced the tensions and changes of the data practices that people use at each site. This included the socio-cultural values that practitioners followed at each site that influenced and justified their data practices, ultimate affecting the meaning ascribed to the data. Additionally, the methodology illuminated the institutional context, public policies, and the legislation that constrained those practices [121].

The *data journey* method uncovers the multiple elements involved in the making and understanding of data. Thus, responding to the call of developing methods that provide a more situated and reflexive approach [25, 26, 23, 19] rather than falling into the reductionist and technocratic epistemologies, which assume that data are objective and neutral [19, 122].

A second approach is the *Thickening of data method* [123]. This method of analysis allows for capturing situated experiences and processes on social media data. *Thickening of data* refers to enhancing the depth of each data point rather than increasing the number of data points. *Thickening* data collected on social media is a form of qualitative analysis that consists of building a multi-layer model in which each layer satisfies a research objective.

The authors of this method initially proposed a three-layer model with the caveat that this is not definitive and insist on adding multiple layers. According to the proposed method, the first layer provides contextual information including the specific technical and social conditions in which the examined practice emerged. The second layer consists of detailed descriptions of the practices that are being examined. Lastly, the third layer consists of capturing the experiences of the people whose practices are the object of study [123].

These theoretical frameworks help recognize the human interventions and social constructs embedded into data [25, 24]. However, what remains understudied is how to surface the role of human intervention and subjectivity within the practices involved in analyzing social media data to draw conclusions and inform NGOs while caring about the context and representation within data. The research I present in this dissertation is a response to the several calls for a more cautious and caring understanding of data, specifically for social media data [19, 18].

3.2.4 Data Ethics

The extensive amount of social media data publicly available not only has changed how we examine human networks and communities, but extracting and interpreting this content has challenged the traditional understanding of respect, beneficence, and justice that protect human subjects in research [16, 124, 125].

When dealing with the data produced by people, the ethical inquiry moves away from the traditional understanding of harm, becoming a more abstract issue obscuring the impact of privacy and data discrimination [124, 126]. Through the use of different methodologies, particularly machine learning, user-generated content has been extensively used to infer new meanings of people's behavior across contexts, which has become a valuable resource for many disciplines. However, the ethical implications of collecting, storing, analyzing, and interpreting user-generated content, while still being defined, suggest the need for rethinking the ethics that have been guiding human subject research so far to incorporate the fact that data are infinitely connectable, repurposable, and updatable [126].

Furthermore, the challenges of regulating the collection and use of user-generated content also derive from the (mis)conceptions of what constitutes human subjects research in the context of big data [127]. For example, current US research regulation exempts research projects that make use of data that are public and available under the assumption that they pose minimal risk. However, as pointed out by Metcalf et al., secondary analyses of publicly available data or combined with other datasets can pose serious risks [126]. Additionally, the nature of user-generated content, of being publicly and massively available, challenges practices that have been used to maintain respect with participants during research, such as informed consent, which is difficult to obtain from people generating social media data. These challenges highlight the asymmetric power dynamics between researchers and online research subjects, because subjects are usually unaware of being monitored, and their lack of both the capability to decide which of their data can be collected and to protect their data [124, 16].

In response to the increasing debate over the ethics of online data research, academics have inquired about the practice among researchers when using social media data for research [124]. Vitak et. al, conducted a survey of 263 online data researchers to identify areas of agreement and disagreement in regards to ethical practices when collecting and analyzing social media data. Their findings show that across disciplines, there are four cohesive research practices when using social media data: (1) removal of individuals when they formally request it, (2) discuss with colleagues and review boards ethical practices, (3) make results available to participants after the completion of the study, (4) being careful on reporting edge cases. The findings of this research also indicated areas of disagreement when collecting social media data, including (1) disregarding Terms Of Service, (2) obtaining informed consent from data subjects, and (3) collecting data from non-representative samples [124]. The disagreement on research practices highlights the emerging challenges of researching online communities imposed due to their complexity when defining context and norms, which can not be immediately solved through the traditional interpretation of the values of respect, beneficence, and justice, as suggested by the Belmont Report [128].

Lastly, in response to these challenges, there have been government initiatives that aim to address the risk of privacy posed by machine learning inferences from different kinds of data. Notably, these initiatives emphasize the protections on the use of data rather than the collection of it. For example, the Obama Administration's Big Data Initiative focused on user-based protections by recommending the increase of technical expertise to address the discriminatory impact of using big data analytics and provide consumers with the necessary tools to empower them in their decisions of use and disclosure of data [125].

CHAPTER 4

STUDY ONE: GRASSROOTS DATA PRACTICES TO ADDRESS PUBLIC SAFETY

4.1 Introduction

Social media have come to empower social movements across the world [54, 129]. One of the shared conditions that prompt people to turn to computing infrastructures is the collapse or inattention of established institutions to address the physical, social, and cultural conditions of oppression or insecurity. It is within this context that I examined the particular conditions in and around Mexico where citizens have turned to Facebook to catalogue data about local crime. Facebook groups allow users to report local incidents with some degree of anonymity, reducing the burden of retaliation from authorities. Due to the Facebook sites' ability to keep their content visible to anyone on the internet by default, these pages also become sources for the production of data and evidence. For many citizens in Mexico, Facebook has become the sole source for learning about issues that are suppressed or ignored by either the government, or by the Mexican media. In this first study, I specifically analyzed the strategies administrators used to turn user generated content about instances of local crime into data to scaffold action among members of the Facebook groups and pages.[1] Additionally, I examined their practices for building online communities and curating data that had been omitted from official database and mainstream media in Mexico.

The purpose of this study was to gain further insight into how local knowledge was being recorded and turned into data so that it might augment or contest incomplete official records [2]. Thus my guiding research question was the following *how do local citizens use the content and features of social media as a resource to understand and organize against*

[1]Henceforth, Facebook sites.

[2]This study was published in CSCW 2018 [28].

local crime?, which I broke down in the following questions:

1. **RQ1:** How have people adopted Facebook groups to collect and exchange information about crimes in high violence contexts?

2. **RQ1a:** What are the different practices that members of these groups adopt to establish online communities?

3. **RQ1b:** What are the users' practices to scaffold the information collected on these online communities?

4. **RQ1c:** What are their challenges to scaffold the information collected?

At the time of conducting this study, my overarching research agenda was to better understand how local citizens and activists use the content and facilities of social media platforms as a resource for organizing toward systemic social change. With respect to this study, I was concerned with gaining further insight into how local knowledge was being recorded and turned into data so that it might augment or contest incomplete official records. While the data in these pages were valuable because they provided an account of officially unrecorded recurring violence from citizens' perspective I was also aware of the limitations of these accounts as sources to augment official crime data. First, the data collected are hard to verify because they cannot be validated through any institutional process. Second, the data are unstructured which creates a barrier for use as official evidence. Third, due to a widespread lack of trust most of the citizens involved in the administration of these pages work anonymously, limiting citizens' ability to link online and offline activities. Despite these limitations, however, I found that administrators have been able to build robust communities that have translated data into actionable results.

In this analysis, I demonstrated how citizens manage and make sense of collective data, while also dealing with safety, privacy and infrastructure constraints. Second, our work is in conversation with other civic domains that benefit from collective data gathering e.g

[92, 130, 91, 66]. However, we examine a more complex space where the safety of the contributors is at stake because of the complex relationship to both social and state organizations involved in creating and perpetuating harm against their communities. While our research focuses on Mexico, the issues and insights presented in this work are not unique to the Mexican experience. In particular, our insights regarding data challenges resonate with other situations similar to Mexico, including contexts coping with sociopolitical turmoil and high levels of distrust in authorities (i.e., governments; policing agencies).

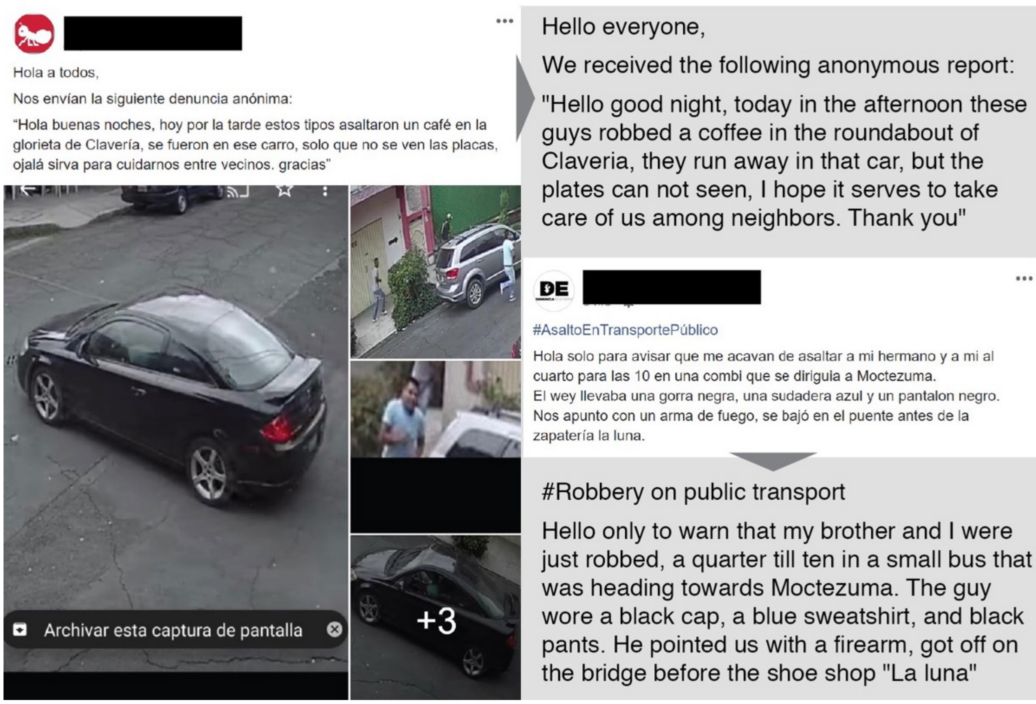


Figure 4.1: Two examples of the kinds of accounts citizens posted to the Facebook sites tracking issues of crime and public safety

4.2 Methods

In order to understand how social media enable collective organization, I conducted a five-month-long qualitative study [3] with citizens that use Facebook to report and track crime and safety issues in Mexico. Through a series of semi-structured interviews, I worked with

[3] During winter 2017 and Spring 2018

administrators of both Facebook sites, which are public by default, and private Facebook groups. Additionally, I interviewed regular users who were active in participating on the sites I identified. The goal of the interviews was to gain an understanding of administrators' motivations and expectations when setting up these pages, as well as citizens' experiences and processes while publishing incidents when they suffered from crime.

Previous conducting the interviews, during 2017, I followed and conducted observations of several Facebook sites that report on crimes from different cities across Mexico, as well as from neighborhoods within Mexico City. Thus, I recruited administrators' participation in the study by approaching the individuals who ran these pages. Then, I continued to increase the sample by seeking additional groups and pages according to three main criteria, a) geographic location, b) content and c) interaction among members. Across these three criteria, I looked for pages whose contributors where within Mexico, for content focused on issues of crime, security, and safety, and for an active base of contributors. Sites outside of Mexico, or those focused on general neighborhood concerns—such as trade, events, classified announcements, road advertisements—or those without active interaction among members were discarded. Likewise, I discarded pages that had a single source of content (often the administrator) because my purpose was to gain more insight into how citizens mobilized user-generated content to translate online collective collaboration of citizens when reaching into offline action. The pages I focused on recruiting to the study contained detailed testimonies of crimes experienced by community members. In some cases, the testimonies included pictures of the offenders, missing or abducted people, or pictures of stolen property (most often stolen vehicles). Videos usually showed criminal assault and theft such as carjacking, kidnappings, burglary, and robbery. Finally, the text of these pages provided exhaustive descriptions of crime, including the address where the crime happened, a physical description of the offenders, and modus operandi (see Figure 4.1).

In total, I contacted forty-five sites; I received a response from ten administrators. Of those ten, I was able to interview seven and three dropped out of the study last-minute due to concerns about safety. For the recruitment of citizens who were either following or had shared information on the Facebook sites I identified, I followed two strategies. First, I recruited citizens through personal networks based on prior work I had done in Mexico. Second, I looked for people who were recently active on the Facebook sites and contacted them directly through Facebook messenger. Across both modes of recruitment, I contacted twenty-five citizens and received responses from seven people.

The concerns about safety carried through the all of the interviews: the participants I spoke with were all very concerned about protecting their identity. This limited the amount of demographic data I could collect. Of the 14 people I interviewed, five were women and nine were men. Participants' ages ranged from 24 to 48. Eight of the participants reported having a bachelor's degree, three reported having master's degree, one participant reported having Doctoral degree, and two participants refused to give this information. Four of the administrators managed Facebook sites dedicated to collecting data focused on the state of Mexico, including the municipalities of *Naucalpan, Ecatepec, and Nezahualcóyotl*. Two administrators focused on Mexico City and one focused on the state of Jalisco. The seven citizens I interviewed reported following Facebook sites based on the location of work and home. All of the participants lived in Mexico City and they followed pages from the municipalities of *Coyoacán, Azcapotzalco, Tlalpan, Álvaro Obregón and Gustavo A. Madero*. These regions illustrate a diversity in crime and economic conditions.

The interviews lasted between 30 and 76 minutes (average of 50 minutes) and were conducted via Skype, Google Hangouts, Telegram, or over the phone. The manner of interview was left to the participant as a way to help manage personal safety and privacy. Participants had real concerns of having their identities exposed through the interviews and I collectively took steps to maintain their anonymity, including using only voice (no video) for interviews. The interviews were audio-recorded and then transcribed in Spanish and translated to English for the purpose of reporting here. In reporting the interview data below, I anonymized the names and genders of respondents to protect their identities.

To analyze the data, I inductively and iteratively coded the transcripts, using memoing and coding. The initial codes focused on understanding general practices of administrators and citizens around data, community, and action; the memos connected these codes with data examined in the Facebook posts. Following a standard practice of qualitative analysis, I then developed a set of themes describing common practices in three areas: within group practices of administrators and citizen, between-group practices as administrators and citizens worked together, and extra-group practices as administrators used the data shared on Facebook to contact outside authorities in order to address the conditions in their communities. These themes form the basis for my understanding of how Facebook enabled a set of datasharing capacities within a number of concerned communities in Mexico City, and how those capacities translated into real-world action to address the conditions of crime and violence being experienced in the city.

4.3 Human Infrastructures: Citizens and Administrators

Administrators of these sites, the people who initiated the Facebook groups, often took on additional work to manage, curate, and share information provided by regular citizens. My analysis surfaced three practices that administrators engage with —community, data, and action practices— to collect, curate, and publish data about public safety that would otherwise go unreported.

Administrators' expectations with these pages were diverse but they all agree d that building online communities where citizens could express their concerns and organize to find solutions to issues of local crime was a priority. For some administrators, the main motivation was to gather and centralize data about current events in their neighborhoods as a way to overcome misinformation and lack of official data. For others the expectation was to maintain an independent record of crime and its effect on their communities. Juan, one of the administrators shared a common view:

In recent years in our neighborhood, the problem of crime and insecurity has

> *grown a lot and we noticed that people distrust so much when they have to file complaints and talk with authorities. [Citizens] prefer to have an anonymous channel where they can publish the security problems that are happening and that is the motivation for creating this platform.* — Juan

Across the different motivations for establishing and maintaining the Facebook sites, administrators played a key role in defining the criteria for filtering the stories and data that were published. Administrators also decided if and how they will take action. It was the administrators' criteria that ha d the largest impact on the quality and type of data posted to the sites. Guiding their community members on what and how to post was essential to translate the data into further action. As Vanessa put it,

> *Our goal is to encourage people to file more complaints [with the police] and if they do not feel safe doing so, we offer them the alternative to do it collectively in these pages.* — Vanessa

The administrators understood that building a record of incidents on Facebook was not enough, that citizens needed to report things to the police as well, and so part of their work was identifying information that would strengthen such reports and encouraging the citizens who visited their pages to make those reports.

Underlying this need to increase the number of reports to the authorities are issues of quality of data. Here I mean the level of detail in these narratives, and the type of data such as videos or pictures. In this study, I identified and organized five categories of the most frequent accounts that were gathered through these pages, including: social complaints, which refer to any citizen report regarding the lack of city services such as water or electricity street crime, which encompasses incidents like street muggings, carjacking, and burglary disappearances, which refers to any report or information of missing people in Mexico this is often connected to kidnapping and hostage taking for ransom community service including citizens' reports on missing dogs, donations, reporting accidents and real time updates of the city (e.g massive traffic); and finally calls to action where citizens encourage their neighbors to file official complaints or go to the police.

Across these different categories of posts and reported content, administrators worked to build a trust with their community. As previously mentioned, distrust, censorship, criminal impunity, and a dearth of accurate information were the main factors that contributed to the birth of these online communities. Therefore, administrators' practices were directed to counteract those factors. In order to build their communities, administrators first needed to attract the attention from inhabitants of the affected neighborhoods. Then, to keep the community growing and to sustain meaningful engagement and the posting and sharing information, they had to ensure that the information gathered was useful and helped the community act on the specific incidents reported.

4.3.1 Community Practices

Administrators' practices to build communities varied depending on the targeted location of the Facebook site. The interviews we conducted revealed three common practices the administrators used to encourage trust and citizen participation. First, the administrators filtered and monitor ed members who ask ed to join the group or follow the page. As mentioned above, we identified both Facebook groups and Facebook pages that focus ed on citizens sharing safety related issues.

There is an important difference in how these sites operate on Facebook: pages are public, meaning anyone can choose to follow a page and will get updates as new posts arrive (to the degree that Facebook propagates these updates to individual timelines); conversely, when creating groups the administrator decides whether to make it publicly available for anyone to join, whether to require approval for members to join or whether to keep it private and make it visible to other Facebook users only by invitation. The choice to setup a page or a group was intentional among the administrators we interviewed. Some administrators preferred to create groups in order to have more control over who became a member.

> *We did have bad experiences, and because of that now what we do is, first verify that the profile is real. So we make sure that the profile was created more than one year ago and also, we verify that the profile has real pictures, not only one or two. We also verify that the profile has more than 200 friends, and with that [information] we make sure that it is a real profile. Then, if we can verify the friends of the profile, we also do that.* — Juan (administrator)

Even after admitting new members, administrators reported continually monitoring member's behavior and interaction in the groups and pages. These practices were put in place to establish the integrity of the group and to maintain community norms to help ensure people treated each other with respect and kept on topic.

> *Generally, we do not monitor all the comments, but when we see that they [citizens] are insulting to each other or something like that, we remove them. We also want to maintain good communication between people preventing them from reaching that kind of thing. But we do not have a very hard policy to moderate or comments unless it is very serious.* — Juan

The second shared practice centered around how administrators curated and shared content to attract attention from a broader audience. Here the administrators took an active role in selecting data to share in posts in order to build a following. Some, like Bruno, used public events that were not specifically focused on crime in order to draw attention:

> *During my last year of college, I helped to distributed toys to the children in my neighborhood and I uploaded the photos of the event [to the page]. And there was a very good response from people [followers and community members of the Facebook site], I think I gathered 30 or 40 more followers that day and then I decided to continue feeding the page with news, with the information people sent me and it was growing very organically ... I did not expect people to start sending me information ...* — Juan

By focusing on positive activities within the neighborhood, Bruno created a site focused on community news that then became a place where others began posting and sharing information Other pages were focused from the beginning on issues of crime and neighborhood challenges.

> *At the beginning what we did was to invite the community to be informed about the security issue So, we uploaded data, information, statistics, but we also shared reforms that had been made to different regulations, police security, traffic regulation, etc. And we uploaded [newspaper] notes about all these issues and people started giving their opinions and discussing these issues.* — Vanessa.

Finally administrators worked to build alliances with local institutions and authorities to strengthen the community so that the information collected on their site would lead to police investigations or otherwise result in some response for the community. Some of this was apparent in the quote from Vanessa above where she pointed to specific instances of reforms as examples of how sharing such information could lead to change. She went on to point out the challenge in getting outside authorities to pay attention.

> *When we started it was very difficult for us to get recognition [with the page], we wanted to get the attention of the neighbors we want them to recognize us [as the people behind the page]. We also try to connect with some people who were leaders, and it has not been easy because the leaders [from government] belong to political parties* — Vanessa.

Building alliances on multiple fronts was a challenge for the administrators, however, given the severity of the conditions in their neighborhoods, each persisted in order to build local coalitions to address the issues of violent and serious crime. What they each recognized was that there was power in the production and sharing of data.

> *We do try to encourage them [citizens] to give us information through the page, and we appreciate that they share with us the information about criminals to alert other people and others can be more careful. Even so, this will not have an effect on the government statistics. They [the government] can argue that crime is being reduced just because people are not filing complaints and that is exactly why we tell them [citizens] that is very important to go to the police station and file the complaints but sometimes they do not even know where the police stations are.* — Jorge

The difference here is that the data collection mechanisms were not explicitly tied to civic bodies as other kinds of systems have been so the administrators had to build out additional practices for working with data and then turning those data into opportunities for action.

4.3.2 Data Practices

Where community practices concerned how the different Facebook sites were run, data practices cover the range of management activities involved in selecting and publishing data. From our interviews, we identified two data management approaches. In the the administrator acts as gatekeeper; they post all of the data. A side effect of this mode of site management is that visitors to the site are unable to identify the source of the post because all the data appears to come from the administrator While this protects the identities of individuals contributing details about crime, it does mean participation in the site is co n strained by the gatekeeping administrator. The second management practice simply allowed site members to freely post stories. In this case, the administrator' role was more akin to that of a moderator and less like gatekeeper.

In the sites were administrators were the gatekeeper, they reported having two main sources of information members of the community or local authorities. The type of data varied depending on the source. When community members reported information, their focus was mostly on the categories of social complaints, street crime, disappearances and community service. In contrast, when the information came from local authorities the content focused on sharing private information and making a call to action. For example, in some cases the authorities sent pictures of people who had been arrested and detained in police stations:

> *There are police officers who let us know when they arrest someone and take him to the police station and send us pictures. We share them [pictures] on the page so if somebody recognize s the [alleged criminals] then the citizens can file a complaint. And there have been cases where citizens recognized them, and they file complaints.* — Felipe

The purpose of sharing these pictures was to encourage citizens to file reports in case they recognize the arrested individuals from other criminal activity. However, these practices raise d some concerns, even among the administrators, since much of this information is hard to verify and affects the safety of the alleged criminals. As a result of these concerns, before publishing any data many administrators filter and curate the content. The former refers to choosing what pieces of information are published and what data are not published. The latter refers to the practices of deciding how to present the data; for example, editing or cropping pictures revising text, and otherwise manipulating posts to focus attention on the details deemed most important.

Most of the participants reported having strategic criteria for deciding what content should be published in order to prevent the propagation of false or outdated information. The goal here was to establish a reliable channel for information that did not expose individuals unnecessarily. For example Vanessa remarked in response to how she managed her site:

> *Then [we] started to filter the information, for example, [citizens] cannot upload [pictures of] faces. We put filters on the face so it is no longer visible.*

[These filters also apply to us] because we may decide to upload the information, but without including the full names and without being too sensationalist.
— Vanessa

Among other aspects administrators considered was the authenticity of the profile of the person who sent the information, the quality of the data the level of detail and the type of data such as photo or video. Beyond following these basic criteria, administrators also chose between stories that better represent ed the tone of their site keeping in mind that each of these sites had their own character. As noted above some grew out of sharing general interest topics in the neighborhood, while others were more specifically focused on issues of crime from their outset.

Across these practices, however, administrators recognized the limitations to how information shared on their sites would help them corroborate with outside institutions and authorities In either case, the data practices whether as gatekeeper or moderator where about establishing community norms, grooming information that was posted to their site, and finding ways to effect change The administrators recognized that this last step, while crucial, was also not something that occurred naturally from simply cataloging incidents of crime on social media.

4.3.3 Action Practices

The communities we examined in this study are built, in part, as result of the data reported by the citizens. Based on these reports administrators intervened and collaborated with other stakeholders to address the specific instances of crime. This, in turn, helped to improve the reputation of the online communities on Facebook, increasing trust and encouraging participation from citizens. Establishing these ties also help ed to legitimate the data gathered among the members.

We describe these interventions as action practices and their outcomes depend ed heavily on the neighborhood, the available data and the amount of time the communities had been working together. As we mentioned above these are location based communities. Although most of the neighborhoods we examined were experiencing similar conditions of violence and lack of safety, the urgency and the scale of the problem varied from wealthy to more marginalized communities. Different neighborhoods also reflected disparate approaches to protecting their communities. Like wise, the kinds of responses they received from local authorities differed along expected lines of socio-economic status.

Translating data into actionable results required administrators to build alliances with key stakeholders. These alliances leveraged the efforts from the data and community practices the combination of both a vibrant and vocal community presence on social media with data about instances amplified the capacity of the neighborhood to receive attention from different stakeholders. Among these stakeholders were government institutions, local authorities and media outlets.

The processes employed to connect to these different stakeholders varied greatly depending on the neighborhood and the degree of anonymity the administrators' sought to maintain. Participants who were afraid of retaliation preferred to keep an anonymous profile (on Facebook, this requires creating and maintaining a pseudonymous profile), and exclusively using online interactions with stakeholders. On the other hand, some administrators reported being comfortable showing their identity while making reports to local authorities and news media. Local authorities such as the police, prosecutors and neighborhood leaders were among the main collaborators with whom the administrators partnered. The situations where they collaborated most were often related to lack of services (e.g lack of water, street lighting) lighting), and not on issues of crime. By partnering with local authorities on lower risk issues, the administrators were able to build relationships that could then be used when addressing the more serious issues of crime and public safety.

The most common pattern of action followed when citizens tried to solve an issue through the proper local or municipal office but due to a lack of action, turned to the online communities to report their concerns and bring other kinds of public pressure to bear. The administrators would then intervene on behalf of the online community members. These interventions were both online and offline, usually navigating through different platforms to amplify their efforts and getting a response to the given issue. Juan provided a clear example of these kinds of practices:

> *What we do is that we skip the municipality office. For example, there have been people who had already gone with them [to complaint] and the off ice gives them a tracking number but they do not take care of the issue. So, what we do is to contact the Urban Management Agency of Mexico City because they are in charge of following up with municipal offices. We contact this Agency via Twitter and give them the tracking numbers of the issues we collect. The Agency pressures the municipal offices and they even sometimes send somebody to solve the problem.* — Juan

These kinds of service issues were possible to ameliorate through social media platforms because administrators could leverage the wide distribution of information that social media supports. As others have observed, social media, in cases like this, can help citizens bypass normal channels of official response [92].

Administrators also built alliances with police officers and prosecutors. Most of these alliances happened online allowing both parties to maintain anonymity. These collaborations, in contrast to the service issues above, were usually initiated by police officers who contacted the Facebook sites with different kinds of information. In some cases, with pictures and details on people detained at the police stations, and in other cases by providing updated information on developing situations. The purpose of sharing pictures of people detained was to encourage members of the online communities to file complaints. Jorge described the overall practice as a way to hold policy makers accountable.

> *We do try to encourage [citizens] to give us information through the page, and we appreciate that they share with us the in formation about criminals to alert*

> *other people and others can be more careful. Even so, this will not have an effect on the government statistics. [The government] can argue that crime is being reduced just because people are not filing complaints and that is exactly why we tell [citizens] that is very important to go to the police station and file the complaints but sometimes they do not even know where the police stations are...* — Jorge

The strategy here was motivated by the large number of unreported crimes and an attempt to the consequences that under reporting has for funding and staffing public safety. While this practice was reported by most of our participants, not everyone approved of it.

Prior work has pointed out the importance of considering the role of trust in the relationship between community and police [131], and how trust affects the usage of In our case, when the police post pictures of alleged criminals it raised concerns among some administrators that such information might lead to false accusations. The balance administrators had to strike was often difficult. On the one hand, connections with the police and prosecutors could lead to more up to date information about current incidents and prompt quicker response from the police (in a context where police response is typically very low); but on the other, administrators had to trust the motivations of the police in sharing information, again in a context where trust in the authorities is low due to a history of complicity in crime. Due to these concerns, some administrators refrain ed from publishing pictures, videos or any material on the Facebook sites that might have led to false accusations. In spite of these challenges, we learned that by voicing and sharing their opinions, citizens were encouraging collective action among police and administrators. This in return contributed to raised awareness among citizens not only to report crimes with the police, but also to demand more accountability from the authorities.

Similar to the findings in recent work in India where an citizens expressed a desire to report crimes via social media [132], the use of these Facebook sites in Mexico not only help to overcome social fears while communicating issues of concern but also help increased the personnel available for identifying crime. In our case, we additionally found that administrators established communications with news media outlets and journalists. Communication was often initiated by administrators who shared relevant local news to journalists hoping they would publish them, but once relationships were established, the sites became an established source:

> *In the beginning, when I felt that it was worth it that was important I reach ed out to [journalist and newspapers] and asked them to please publish the case. But, lately even if I do not contact them, [journalists] take the cases, the information that I publish.* — Bruno

Some of the success administrators had when reaching journalists was due to having a background in journalism and knowing which media outlet and journalist to contact.In one particular example, an administrator posted about a missing girl that resulted in widespread social action:

> *One night I received a photo of a girl who disappeared, so all I did was post it on the group I manage the next day the case got more attention because a journalist contacted the victim's family in the middle of the night, interviewed the parents, made a video and posted it in the newspaper website and the story became viral. Then, people started creating Twitter accounts demanding justice for her, and then a very particular phenomenon happened because people began holding meetings that were not even organized by the municipal government but by students, and people from the neighborhood they also organized marches, and every thing happened only because of the call that was made across different accounts in social networks.* — Bruno

While the outcome of this event was positive because citizens were able to organize and draw enough attention from the government to collaborate and help the family of the victim the event also illustrates how contingent this alliance is due to the unequal relationship between administrators and journalists. Ultimately, administrators depend on their network, their reputation and the power of their allies to scaffold action through the data collected on their online communities.

4.4 Implications: From the Individual to the Systemic

One of the features that stood out in analyzing how administrator and citizens used the different Facebook sites to collect and share crime data was that their strategies were case-based. While the data in these sites was valuable, their case by case nature meant that administrators and authorities could only ever address issues individually. The challenge here is that the kinds of crime that were affecting the communities of Mexico City was not the result of small, acute acts by a few individuals, but was systemic from the presence and power of the cartels, to the complicity of local authorities and so any meaningful response to these issues would also need to be systemic.

The interviews revealed some of the reasons these communities rose up to respond to issues in a highly localized way. First, the quality of data, which is usually unstructured and hard to verify, meant that the ability to sort and make sense of incoming posts, images, and videos was bound by the work it took for administrators to go through the content and curate it according to their local knowledge. In some cases, administrators reported receiving an overwhelming number of cases each day. This not only complicated the tasks of filtering and curating data but also the process of making sense of problems at a larger scale. The second difficulty is that the collaborations with outside stakeholders and institutions were unstable which meant administrators had to build and maintain those relations at the same time they were sorting through incoming posts. Finally, members' participation was often sporadic, so the number of individuals posting information or contributing to the Facebook site would wax and wane as conditions within the neighborhood became more strained or improved there was not a stable set of contributors that could be tapped to help share the burden of managing site data or working with outside institutions.

Despite these limitations, the content in these online communities could have a significant role to play in Mexico due to the diversity of crimes, testimonies, and data formats that are concentrated on these sites. Additionally, these data, if exploited, could help to identify patterns and more concretely identify the scale and type of crimes. However, to amplify administrators and contributors' efforts, these need to transition from individual to systemic practices. Much in the way that Tufekci discussed how the intersection of on the ground organizing in combination with effective social media use can raise the capacities of small grass roots organizations [54], I would argue that a similar complementary apparatus is needed to more fully empower local communities when confronting systemic issues.

4.5 Conclusion

I conducted this study in 2018, and I recognize that at the time, I assumed that the existence of crowdsourced data reporting on local crime was valuable enough and that it was a matter of developing the right tools to transform such data into actionable knowledge. However, this is a much larger endeavor than the mere creation of tools; it entails recognizing data as the result of social relations with a unique political, cultural, and historical context of production and circulation, as well as examining the larger ecosystem in which we intend to instrumentalize such data [92, 118]. To define what it means to leverage the content from social media data and develop the mechanisms to do it are the central threads that weaves this dissertation together.

In the following chapter, I describe an initial effort to systematize the process of quantifying, interpreting, and situating social media data as an alternative to monitoring the ongoing crisis of disappearances in Mexico. This study is where I begin to pay much more attention to the role of information infrastructures in defining the characteristics of social media data that could make it count as evidence, which restricts how to render this content usable and actionable.

CHAPTER 5
STUDY TWO: DATA MIGRATIONS: USE OF SOCIAL MEDIA DATA AS EVIDENCE

This research study represents a first step in examining the interplay of the formal and informal data practices of communities, civil organizations, and government entities in the implications of mobilizing social media data into institutional contexts.

In this study, I followed an iterative human-in-the-loop computational approach to explore whether citizen reports of abductions concentrated on Facebook groups can provide evidence to complete official records on the ongoing crisis of disappearances in Mexico. The contribution of this study consisted of mobilizing social media data to an offline context by conceptualizing three essential practices –finding, scaling, and verifying signals– and identifying seven data characteristics that rendered this mobilization possible. The second contribution consisted of establishing a data gap between the number of cases officially recorded in the government database of missing people compared to accounts of missing people circulating on Facebook.[1]

5.1 Introduction

The ubiquity and embeddedness of user-generated content and local expertise clustered across social platforms create a potential for aiding the discovery and monitoring of systemic issues that are difficult to track through traditional media mechanisms due to stigma or institutional reluctance [133]. Examples include tracing incidents of mental health crises on college campuses [102], or, in this case, tracking an ongoing human rights crises. Building on exiting work, we know that social media platforms can be turned to for evidence to hold institutions accountable [56, 134, 135]. An example of this practice are the hundreds

[1]This study was published in CSCW 2020 [29].

of YouTube videos documenting human rights violations in the civil war in Syria [136, 137, 138]. Similarly, social media data has been useful in identifying cases of violence against specific populations in the absence of official records [139, 140], and in extending the categories of data collected by governments during humanitarian crises [141, 142]. While user-generated content ought to be a valuable resource for investigations, it remains unclear how to apply social media data as robust evidence, particularly in instances where non-government organizations are seeking legal interventions. Furthermore, the existing research on these topics is limited to examining eyewitness videos and images shared during protests; there is less work on examining the non-video content that could be migrated and leveraged by NGOs.

Across the different applications of turning to social media data to understand human rights crises, a consistent challenge is preserving meaning as content is taken out of context: without that context the documented experiences lose meaning and value [16, 56]. I refer to this as the challenge of *data migration*, since drawing insights and leveraging social media datasets requires us to *migrate* data from online communities where the data are being produced, into other contexts where those data can be operationalized into actionable insights and evidence. The challenges that *data migration* entails are not new. Previous research 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 production [30, 118]. Among the most prominent challenges are: how to make data legible to different stakeholders and "rendered transportable and intelligible" [30], while maintaining data quality, and addressing issues of documentation and provenance. Building on prior definitions of data re-use, *data migration* "implies the communication of something to a set of potentially unknown and unknowable others" [30], while additionally calling out the challenges of moving across technological and institutional infrastructures that were never meant to inter-operate. Ultimately, these challenges indicate that knowledge and insights from social media data cannot simply be extracted from the context where they were produced, to be re-used in a different context claiming a sense of objectivity and accuracy. Instead, they must retain context by addressing issues of documentation, provenance, and recognizing that data are not self-contained units. Therefore they always needed complementary external information to be understood [30]. These persistent challenges suggest that there is still a need to develop mechanisms that allow us to manage and retain context as we migrate data from online communities to guide offline action. This study aims to address these challenges by focusing on the issues of documentation, context, and provenance to facilitate data migration from Facebook groups to inform NGOs' confronting an on-going human rights crisis in Mexico.

In this work, I examined reports of missing persons in Mexico circulating via Facebook groups. Through an iterative process, I identified seven data characteristics that enable mobilizing this content and establish a data gap between the number of cases officially recorded in the government database of missing people as compared to accounts of missing people shared on Facebook.

I focused on missing person reports because there has been an increasing number of missing people since the most recent war on drugs that began in Mexico in 2006 [143, 31, 39]. Several national and international organizations have urged the Mexican State to investigate and take action due to the increasing number of missing people and clandestine graves [144, 145, 146, 147]. Despite the attention and scrutiny from the global human-rights community, the exact number of missing people remains unknown due to inconsistencies between the figures provided by government reports and those provided by national and international NGOs [148, 149, 150, 151, 152]. The situation has only become more urgent as recent discoveries of clandestine mass graves with unidentified bodies throughout the country have come to light [152, 153, 154, 155, 156]. As a consequence of the crisis, a considerable number of NGOs and organizations supported by relatives of missing people emerged across the country to ask the government for justice and a resolution to the disappearances across Mexico [157, 158]. One of the most notable is the collective called

Movement for our disappeared in Mexico[2], which is composed of more than 60 NGOs from 21 states. These NGOs are responsible for finding a large number of clandestine mass graves [152, 159, 160], and they have successfully collaborated with the government to craft the General Law on Forced Disappearance of Persons which stipulates responsibilities and procedures for recording missing persons [161, 162, 163, 160]. The role of these NGOs in the ongoing crisis is crucial due to the absence of experts and tools for forensic identification, the lack of knowledge of the context in which disappearances take place, and the limited resources to conduct searches on behalf of the government [158].

The combined scarcity of data, resources, and expertise for finding and maintaining evidence of the on-going crisis also suggest opportunities for computational approaches that seek to identify and aggregate data from novel sources. The challenge here is three-fold: first, identify signals in social media data streams that can be linked with official data records; second, establish the scale at which particular issues operate in social media; and third, verify the quality of social media data streams. Addressing these needs requires developing a robust method of collecting, analyzing, and cataloging social media born data to be used as evidence in the legal and policy context of Mexico. These challenges are deeply human-centered and place-based, and build on prior empirical research that established data-work practices among NGOs and the cooperative data-collection practices deployed by Facebook group administrators and users [28].

The guiding research questions of this study were the following:

1. **RQ 2:** How can meaningful information be identified on social media platforms for use as evidence by civil and non-profit organizations seeking legal and community interventions?

   (a) **RQ 2a:** What are the necessary data attributes of social media data streams to be considered as evidence?

[2] In Spanish: Movimiento por Nuestros Desaparecidos en México, website: https://movndmx.org/

(b) **RQ 2b:** What are the practices involved in identifying and migrating social media data to be used as robust evidence, particularly in instances where NGOs are seeking legal interventions?

(c) **RQ 2c:** How to address the volume challenge of finding evidence across social media data?

(d) **RQ 2d:** How do we establish the scale at which particular issues operate in social media?

(e) **RQ 2e:** How do we verify the quality of social media data streams?

5.2 Context & Methods

Understanding if and how social media data can be effectively mobilized to confront the human rights crisis in Mexico starts with understanding both how NGOs in Mexico are confronting violence and how citizens are using existing social media platforms to collect and share information about local crime.

Building on prior work that I have conducted, I turned to active Facebook groups as a site where local residents organize to track and share information about a range of public safety concerns, including missing persons [28]. Because of how important Facebook groups are to local, grass-roots efforts to track violent crimes, we want to specifically examine how to leverage that crowdsourced content in order to bridge from the locally-focused data collection and sharing to larger institutional responses being advanced by NGOs.

I argue that it is not enough to simply extract data from social media platforms; instead, that extraction and concomitant analysis must be rooted in the standards of evidence and work practices of organizations, institutions, and other stakeholders who could benefit from the findings and insights those data may provide. The methodology that I followed along with my collaborators to bring local knowledge produced online into the offline context was guided by the needs and challenges that NGOs face when gathering evidence of human right violations [36].

The process of this study consisted on blending three stages of qualitative methodologies and computational methods into a recursive pipeline that helped establish the context of the data I were seeking, the scale at which it was present, and finally, the veracity of the evidence that emerged from the data analysis. At each stage, there was a feedback loop where new insights were tested and refined against prior insights, resulting in a robust dataset derived from a social media data stream that is otherwise noisy and imprecise. I develop this approach over three iterations, alternating between establishing the context of our information in the online and offline communities, and adapting computational methods to identify a signal and, once identified, build scale and evaluate their veracity. In presenting this method and results, I refer to online data as any data coming from the social media steam, and to official data as any data produced by governments or NGOs.

I focused on information gathered on Facebook because it is the most commonly used social media platform for learning about crime in Mexico [164, 165]. According to the National Urban Public Security Survey (ENSU by its acronym in Spanish)[3], since the first trimester of 2018, Facebook was the third most used source for learning about public safety, drug trafficking, and crime in Mexican cities [164]. The top two sources of information were televised news and word of mouth with family and friends [164]. Additionally, the use of Facebook to learn about crime has been steadily increasing: in March 2018, 46.1% of the population were using Facebook; by September 2019, 50.9% of the population turned to Facebook as their primary source of information about crime [164, 165]. It is for these reasons we focus on the Facebook data stream instead of Twitter, as is common in other studies.

Examining data from Facebook groups poses challenges of access, searchability, temporality, and virality. Due to privacy options, Facebook groups can be secret, closed, or public. This, in turn, affects the capacity of gathering data, since the privacy settings regulate the searchability of the group. Secret groups are hidden in search, while closed and

[3]In Spanish: Encuesta Nacional de Seguridad Pública Urbana (ENSU). Survey conducted quarterly by the Mexican government that provides insight on the perception of public safety among urban adults.

public groups are visible in search though their content may not be available. Furthermore, the ephemeral condition of Facebook groups complicates data collection, preservation, and verification since a given group can be deleted by group administrators at any moment: deleting a Facebook group is an irreversible action that removes the content from the platform. Lastly, the virality of the platform takes a different form. Unlike in Twitter where both the broadcast nature of the platform and the use of hashtags enable tracing messages through different networks of people, on Facebook, there are no external mechanisms to trace or follow how messages travel across groups, making it impossible to identify where posts start and how they move through different communities on the platform.

An Ecosystem of Data and Databases

In order to begin to understand Mexico's disappearance crisis, it is important to examine how these incidents are being measured and classified. Within the country, there is a complex ecosystem of institutions and databases cataloging missing people: there are at least nine different official records produced by the government of Mexico and many more operated by NGOs and outside organizations. The majority of these databases and tools were intended to be implemented across the country; in practice however, there are wide disparities on how and where these tools are used and enforced with records managed by federal, state-level, or independent organization [39, 148]. Further, each of national databases vary on the source of their information, the purpose of the registry, and the level of accessibility to the public [39]. Since each database has a different definition and approach for classifying who has gone missing, they all reflect a different number of victims [149, 148]. Consequently, there is no canonical record of missing persons, nor is there a reliable way to assemble one from the many different registries.

At the time of conducting this research, there were two national registries of missing persons: The National Registry of Lost or Missing Persons (RNPED by its acronym in

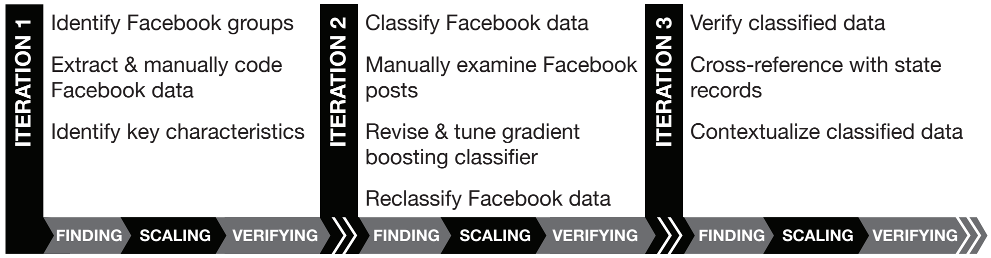


Figure 5.1: Key practices we develop at each iteration

Spanish)[4], which was the first official registry that coordinated across federal agency [148, 39]; and the Ante Mortem/Post Mortem (AM/PM) registry, a similar cross-agency tool for managing information on missing people and human remains designed by the International Committee of the Red Cross (ICRC) [166]. In addition to these two national registries, there are three official databases coordinated by independent government agencies that record cases of enforced disappearances and missing people. However, none of these databases are accessible to the public [148]. This last point is important because even as the Mexican government has attempted to develop data resources to understand the scope of human rights violations, there remains a large number of un- and under-reported crimes. According to a 2017 report, for every crime that exists in the official crime statistics, there are at least nine that were not recorded [167].

Integrating Qualitative and Computational Approaches

To find evidence of missing persons catalogued on social media, we stepped through three iterations of our process. Each iteration was composed of three key practices: finding signals, scaling signals, verifying signals (see Figure 5.1). Each iteration comprised a mixed-methods approach where we used computational methods to gather data, and qualitative methods that helped us contextualize, seek scale and demonstrate the veracity of the data. In the first iteration, we identified the online communities from which we would extract our initial data. Then, we applied computational methods to classify the extracted data. Finally,

[4] In Spanish:Registro Nacional de Datos de Personas Extraviadas o Desaparecidas.

we manually reviewed those extracted data to understand the context of each post. Based on the findings of the first iteration, we refined the computational method, expanded our qualitative analysis, and reworked our data process to identify and collect cases of missing people. In the final iteration, we first confirmed that all of the posts we gathered documented a missing person and then removed duplicate records and verified the final corpus against the RNPED database.

First Iteration The first iteration was guided by our previous fieldwork that explored data practices in local communities and NGOs addressing the current crises of human rights violations in Mexico [36]. This guided us to existing online communities with an understanding of how they operate, and the language and topics discussed within. Additionally, we had expertise working with government databases, which helped us understand the baseline for how the country registered and tracked violent crime. Together, this helped us identify the online communities where we could start collecting data and the topics that were relevant in both the online and offline contexts. We decided to focus on the State of Mexico because according to the RNPED in 2017, it was the state with the second most reports of missing people across the country – 3,890 cases. Using the RNPED, we also identified the municipalities within the State of Mexico that had the highest rates of reported abductions during 2017: Toluca with 438, Nezahualcoyotl with 413, Ecatepec with 408, and Chimalhuacan with 216.

Relying on our prior research on the use of Facebook groups in tracking local crime [28], we were able to identify 45 Facebook groups targeting those four municipalities. Specifically using the names of the municipalities and common keywords used on Facebook groups that refer to tracking crime and organizing against violence (e.g., complaint, neighbors, county)6. Then, using NodeXL [168], we downloaded 13,289 posts and 2,481 associated comments. The posts collected and analyzed in this research correspond to the period of January 2017 to December of 2017 so we could compare with the latest available official data on missing persons.

After identifying online communities, and extracting the data, we built a training set by randomly selecting a subset of 500 posts for hand-labeling. These posts, which were in Spanish, were handlabeled by two native Spanish speakers from Mexico to draw from their expertise to identify slang and colloquial phrasing contained in the posts. Using the hand-labeled training set, we built a Multilayer Perceptron classifier with 40 features, removing image-related features for the classification process. This classifier predicted 851 posts as abductions. Then, two human labelers manually reviewed each of the 851 posts to generate more gold labels and determine the precision of the model. As a result of this process, the human labelers found that only 298 posts were correctly labeled as abductions, giving the initial analysis a low precision value of 35%. Since we were interested in non-video content, we built the first classifier using keywords and text, but that turned out to be the wrong media since people were using images when exchanging information about missing people. We got this insight by manually reviewing each post, and it was only after understanding how people were trading information in these networks, that we were able to implement a more suitable method: an image classifier.

Second Iteration Since the majority of posts concerning abductions included images, we determined that the initial classifier performed poorly because we chose to exclude images. For the next iteration, we used an image classifier to account for the use of visual media when reporting cases of abductions. To better understand what kinds of images were being included in the posts we manually examined each of the 298 posts that were correctly identified by the initial classifier. From this, we identified two different kinds of Facebook posts on abductions: official and unofficial posts. Official posts included images of a state-issued missing person flier that is generated when a complaint is filed with the authorities. These fliers usually include a picture, name, age, physical characteristics, and information about the place where the missing person was last seen (see Figure 5.2). Unofficial posts consisted of homemade fliers that included text descriptions with or without pictures, or homemade videos describing the circumstances of a missing person (see Figure 5.3). Of the 298 posts that were correctly identified by the classifier from the first iteration, we manually identified 155 posts contained an image of a state-issued flier.

We then re-processed the raw post data to scrape and store image data along with the textual content in the posts and comments. In doing this we realized that the official posts always used the same template, though image quality, size, and lighting varied significantly – some images of the fliers were rotated, cut off, poorly photographed or the paper was wrinkled or dirty. We constructed a feature set that derived multiple representations from the post's image. The set included a histogram of oriented gradients (HOG) representation of the images for edge detection, a histogram of values in each color channel (using 20 bins), and word counts of text extracted via optical character recognition (OCR). The intuition behind this featurization was that photos of the official fliers tended to have a large amount of white, as well as edges in stable locations, along with boilerplate text. The initial set of 414 features was reduced to a final feature set of 30 using recursive feature elimination. No dimensionality reduction other than the above was performed. The final model consisted of a gradient boosting classifier using 200 estimators and a maximum depth of 3. The model was evaluated using K-Fold cross validation (K=10). The hand-labeled data, which comprised the training set, consisted of 892 "false" samples (unofficial fliers or another kind of post) and 154 "true" samples (state-issued fliers). Across 10 folds, the classifier achieved a mean accuracy of 95.2%, a mean recall of 76%, and a mean precision of 90.6%. We also implemented a convolutional neural network as a model, using images directly with no manual featurization. It consisted of 3 sequential convolution layers, followed by activation and max pooling, with two fully connected layers at the end. This model was run over 50 epochs and produced a 94.4% accuracy on a train/test split of 66/33. This model was abandoned because it produced similar performance but was much more computationally expensive.

Third Iteration Our final model identified 450 official posts of abductions. We validated each of these official posts following a qualitative approach with a manual review where we confirmed the presence of a state-issued flier, and extracted the name, gender, age, the date, and place where each person went missing. Using that information, we were able to cross-reference each post with the RNPED database. By confirming that each post contained a state-issued flier, we bridged the online with the offline context, linking official reports or missing persons with online information sharing about those cases. Additionally, from our previous work, we understood the significance of the information that accompanies each Facebook post, such as the name of the Facebook group where the post was published, the date, the author of the post, the comments of members of the community, and any discussion that developed [28]. Collectively, this information provided details, such as newspaper articles and other supporting material that further helped us establish the validity of each record. At each of these iterations, the role of the classifier was primarily to identify the media that people use to report abductions. As we moved forward with each iteration, the refinement of the classifier was informed by our previous fieldwork and the qualitative analysis of the previous iteration's output. We documented our process, which later allowed us to identify the insights gained at each iteration, and to further reflect on the interplay of qualitative and computational methods that enabled data migration.

5.3 Findings

Data migration requires a deep understanding of the context where user-generated content is being produced and the needs of stakeholders who may gain insight from that data. In the account I present here, data availability and format were constrained both by the particularities of how citizens in Mexico share information about their missing relatives on social media, as well as what counts as evidence based on the local policy guidelines in Mexico [169].

Through the human-in-the-loop computational approach, I identified and confirmed 484

Table 5.1: Definitions of the characteristics and the stages where they are used

Characteristic	Definition	Relevance	Stage
Official	The data originated from the govern-ment.	This attribute connects the online with the offline context and to enable data migration.	Finding Signals Verifying Signals
Redundant	Duplicate data	This attribute confirms the information and provided us with new insights.	Finding Signals Scaling Signals
Descriptive	Details embed-ded within each piece of data. E.g. source, author, and comments.	This attribute situates the content, indi-cating the online community, the context where it was posted, date, etc.	Finding Signals
Retrievable	Data that has the potential to be extracted or re-covered for further analysis.	Essential attribute to extract and count data.	Scaling Signals
Quantifiable	Data that can be counted	Attribute that measure the scale of the signal.	Scaling Signals
Situated	Attribute related to the location where it was produced	This attribute reveals the location where the data was produced and how it travels across online communities.	Finding Signals
Legible	Refers to the clar-ity of information encoded	Depending on the degree of legibility is the usefulness of data.	Finding Signals Verifying Signals

posts of missing people across the 45 Facebook pages I examined. Each post included an image that was analyzed computationally while any information like comments or the name of the Facebook group to which the post belonged were analyzed qualitatively by visiting the Facebook groups. This analysis focused on identifying which characteristics of these data contribute or limit the process of systematically migrating local knowledge produced online into contexts where these data might be operationalized in a legal or advocacy framework.

Through this analysis, I developed three key practices: *finding signals*, *scaling signals*, and *verifying signals*, and seven characteristics describing the attributes of data within each practice: *Official, Redundant, Descriptive, Retrievable, Quantifiable, Situated, and Legible*. A detailed description of each characteristic can be found in Table 5.1. The seven qualitative characteristics developed from each of the three iterations. This process was guided by our knowledge of the legislation from Mexico and the techniques and methodological

Table 5.2: Characteristics and methods used at each stage

Stage	Characteristic	Method
Finding Signals	Official, Descriptive, Legible, Situated	Qualitative
Scaling Signals	Redundant, Descriptive, Quantifiable, Retrievable	Computational Methods
Verifying Signals	Official, Descriptive, Legible, Quantifiable, Retrievable	Qualitative Methods and Computational Methods

strategies NGOs have developed to identify the probative value of different types of evidence. Based on these elements, I looked for data characteristics that indicated a person went missing, the attributes that allowed to maintain the context of the data as I migrated them, and how they contributed to the operational definition of evidence I outlined above. I describe each characteristic in the following section, linking them with the practices I consider necessary for migrating local knowledge from online communities into the offline context. Table 5.2 shows a concise description of the data characteristics and the methods relevant for each key practice of our process.

Finding Signals

In examining how to migrate social media data streams to the offline context of human rights advocacy, the first question I asked was *how do we identify signals in social media data streams that can be linked with official data records?* By asking this question, I aimed to identify whether or not there are social media data streams that document incidents of missing people. During the qualitative analysis, I found that a key characteristic for identifying content in the social media data stream was whether or not a given post could be described as *official*. Posts bearing the characteristic of *official* are those where the data originated from the government. In this case, posts that were accompanied by images of state-issued missing-person fliers were categorized as "official posts." An example of the official template can be seen in Figure 5.2.

From the 484 posts that I identified, 333 included an image of a state-issued flier, while 151 included an image of a homemade flier. Distinguishing *official* posts helped to identify data signaling the existence of an legitimate missing person claim, one that I would be able to verify in the national data registries of missing people (RNPED database). I also made a baseline assumption that official posts could more readily migrated from the context of the social media data stream because they were derived from existing data present in state databases – they could be cross-referenced and linked to supporting material and meta-data in order to bolster them as evidence.

Scaling Signals

After finding signals that link social data streams with official data records, the next practice consisted of *scaling signals*. At this stage, the goal was to address the volume challenge mentioned in the related work: I needed to establish the scale at which reports of missing people occur in social media data. Scaling signals helped to determine the extent to which Facebook groups are used to distribute cases of missing people. In this case, 30 Facebook groups concentrated all the posts that included a state-issued flier. In scaling signals, I identified that the characteristics of *retrievable, quantifiable*, and *situated* were crucial in enabling us to apply computational methods to collect and establish the volume and source of data. The characteristic *retrievable* refers to the fact that posts from Facebook groups published at a determined point of time can be recovered for further analysis. However, this is a transient characteristic because the existence of the Facebook posts strictly depends on

Figure 5.2: Example of a state-issued flier of a missing person

the continuity of the Facebook communities. While conducting this research, one of the Facebook groups that I followed during the last three years, and from which data was collected, suddenly closed. Although there are methods to extract, collect, and store the images contained in a post to prevent the loss of data after a group has been closed, doing so eliminates other contextual details such as comments and the identity of who created the post.

Within the data we collected, each post portrayed a case of a missing person, and due to the format of the state-issued flier these posts are *quantifiable*, allowing us to identify the number of abductions that were reported within the online communities. In the process of scaling signals, I identified 333 Facebook posts that included an image of a state-issued flier. After removing the instances with duplicate names, we ended up with 308 posts of missing people from the Facebook groups. From the 308 posts, 200 posts reported missing women and 108 posts reported on missing men.

When conducting our data collection, I assumed that people who went missing in a certain geographic location were going to be more likely to be reported in a Facebook group targeting that same location – this is what was captured with the characteristic *situated*. Contrary to the expectations, findings indicated that out of the 308 official posts, only 123 cases belonged to the municipalities covered by the selected Facebook groups: the rest of the cases were located outside that narrow geographic focus. Specifically, 114 cases were of people who went missing in other municipalities in the State of Mexico, 59 cases were of people who went missing in one of the municipalities of Mexico City, 8 cases were of people who went missing from other states across Mexico. Lastly, there were 4 cases from which we could not identify where the person disappeared because the image was not legible.

Each of these three characteristics helps to evaluate the scale at which the cases of missing people operate within social media, which was the second research question. These three attributes allow us to identify, quantify, and document cases of abductions from di-

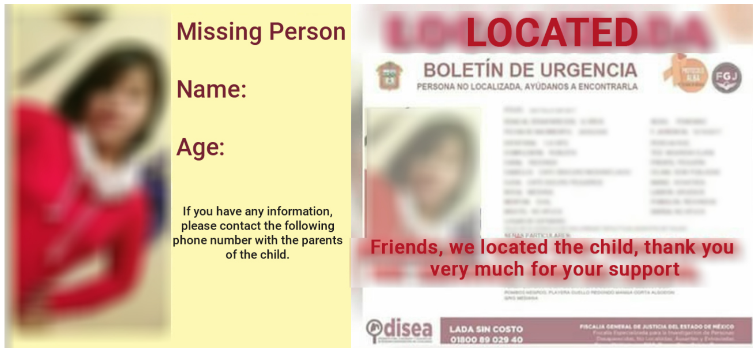


Figure 5.3: Example of redundant information from an unofficial and a state-issued flier

verse online communities at specific times to then compare with the number of abductions reported in official databases.

Verifying Signals

After finding signals that linked social data streams with official data records and establishing the scale at which incidents of missing people operate in social media, I asked *how do we verify the quality of social media data streams?*. I identified three characteristics that contributed to the verification of each instance of a missing person: *redundant, descriptive* and *legible*.

While conducting the qualitative analysis of the 484 posts of missing people identified in the Facebook groups, I noticed *redundant* data either with duplicate names or multiple official notices of the same missing person. Duplicates occurred when the classifier identified multiple posts that reflected the same missing person. Figure 5.3 illustrates one of the 22 repeated cases we identified among the 333 posts that included a state-issued flier. These 22 state-issued fliers were shared across various Facebook groups, at different periods of time, and accompanied by comments and descriptions of the missing people. Rather than perceiving *redundancy* as a deficiency of the classifier, having redundant posts helped us verify the details of each case.

The information embedded within each post, such as the source, author, and comments, composed the *descriptive* attribute. I identified three primary sources of posts: those pub-

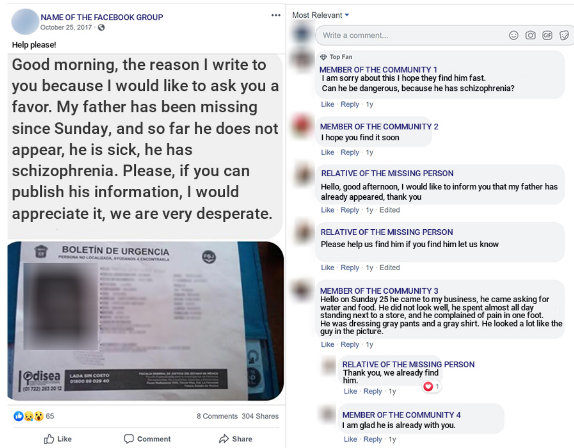


Figure 5.4: Example of the descriptive attribute

lished by the relatives of the missing person, those published by members of local communities, and those shared by the police. Sources of the posts are an indicator of the context and conditions in which the people went missing. Additionally, the comments and the accompanying text of each post provide insights into the particularities of their situation. For example, comments reflect conversations over time among members of the online communities that provide valuable additional information about the missing person. In other cases, the police provided updates or outcomes of the missing person, announcing that the person was found either alive or dead. Occasionally, the information provided in the comments was so rich and detailed that I was able to corroborate cases of missing people with reports in newspapers. In other cases, I found videos and websites in the comments providing more details about the people who had gone missing and about the strategies used by members of the online communities to search for those individuals. As the comments show in Figure 5.4, a man went missing, and his relative provided information on his medical condition to the Facebook community. Also, other members of the Facebook community shared information about the whereabouts of the missing men. The instances where these discussions developed were less frequent, but when they happened, they became rich metadata describing details that would otherwise be hard to find.

In the context of this research, I use the term *legible* to describe the clarity of information encoded in state-issued fliers within the Facebook posts. As shown in Figure 5.2, each state-issued flier includes details of the physical description of the person at the moment they went missing. These fliers signal the existence of an official complaint, but not all of the fliers were clear enough to extract the demographics details of the missing person. The demographic information of each victim is needed in order to search for them in the official database and make an accurate comparison between social media data streams and official data.

From the 308 official posts, I extracted the demographic information from all but five posts that did not include a name (two were abandoned babies, two were children, and one was an elderly person). I then extracted the name, gender, age of the victim, and location of abduction from the 303 official posts and cross-referenced these details in the RNPED database (the national registry of missing persons). I found 27 names in the database; 276 names were not present in the RNPED database despite the existence of an official complaint. The fact that I only found 27 of 303 officially reported missing people confirms the inconsistencies among different sources about the number of missing people [148, 149, 150, 151, 152].

Situating Our Findings

The data migration process not only consisted of deploying a computational model to extract pieces of data from social media platforms, it also included developing a procedure that tied the interests and constraints of those who produce the data to the stakeholders who could leverage that user-generated content. As mentioned before in *verifying signals*, from the 484 posts that we identified, 333 included an image of a state-issued flier, which guarantees the existence of an official complaint, legitimizing the information and linking the online with the offline context. The significance of identifying and retrieving 333 state-issued fliers, of which only 27 individuals were confirmed in the national database becomes clear once we understand the process by which a person becomes a data point in the bureaucratic context of Mexico.

When a person goes missing, the legislation in Mexico has established a protocol to speed up the search and contextualize the conditions of the abduction [39]. The first step is for the family to file an official complaint with the police and establish if the case is related to domestic violence, enforced disappearance, or kidnapping. Then, the appropriate government agencies initiate the search for the missing person. After filing a complaint, there is no clear protocol for the authorities to keep family members informed about updates on the case [39].

Although there is a protocol established to record and search for missing people, through our research, we found that the police usually enforce practices that prevent recording cases of abductions and encourage the removal of names of missing people [148]. Some of these practices include misclassification of the abduction, denial of the filed complaint, and forcing citizens to wait more than 72 hours before filing an official report (although the law indicates otherwise [39]). When the police enforce these practices, the names of those who are missing often do not even make it to the point of having an official complaint, which omits them from any state-run database [160].

For those cases where there is an official complaint, and the names reach a government database, we identified two reasons names could be removed. First, if the police believe that the person who has gone missing was involved in criminal activity, the name is removed from the database and is redirected to a criminal investigation. The second reason results from political change where people are removed when a new administration comes to power and restructures how missing people are counted [50]. In the second case, outside observers have noted that names had been consistently erased from national registries without offering an explanation to victims' relatives [170, 50].

What is crucial to understand from this process is that there is an extended lack of transparency. Relatives of missing people have no recourse to demand accountability from the authorities. When cases are recorded correctly in the official database, there are no mechanisms for relatives to follow up or to know if the names were removed from the database [148, 39]. This lack of transparency in how and why names are removed from the databases complicates the process of situating our findings within the larger context of missing persons in Mexico while simultaneously reinforcing the assessment of disappearances as a present human right crisis. Our results rely on the fact that we have documentation of state-issued fliers recording missing people – fliers that are only generated based on an official complaint which gives us a mechanism for testing the official database. The ability to track and catalogue these reports independently becomes an important tool for NGOs working to address the larger issue.

5.4 Implications: Platforms, Challenges, and Methods Limitations

The data migration process I present here is deeply influenced by the information infrastructures in place. These infrastructures – implemented through formal and informal data practices and linked to social, cultural, and institutional norms [20, 42] – define what counted as evidence, and they enforce an understanding of the scale of the problem. As Bowker observed, information infrastructures are as much social as political, and we need to look at the traces of their history and constitution to understand "what kind of a tool they are, what work they do, and whose voice appears in their unfolding narrative" [42]. Following Bowker's lead, I examined the historical development and evolution of the many different informational infrastructures that the Mexican government has created to record the number of missing people. Through this analysis, we better situate our findings and identify limitations of our approach.

Information Infrastructures

The multiple databases and government agencies that exist in Mexico inscribe the work that has been done by the government and both enable and hinder the coordination among multiple agencies. On examining these multiple entities, I identified the elements taken into account to create these databases, their implementation, and utility. Further analysis of the diversity of databases reflects how the government has understood the crisis of missing people in the country, so far revealing tensions between local and state entities when collecting and classifying data. Ultimately, these databases operate as information infrastructures that enact legal and power disparities between the government and relatives of missing people.

In addition to the numerous state databases, other factors operate as information infrastructures and inform the *data migration* I followed, included police practices when deciding what counts as evidence. Specifically, the sole reliance on police reports to identify missing persons, the criteria that a missing person cannot have a criminal record, and uneven generation and acceptance of reports by local officials. By understanding these factors, I was able to establish the veracity of data pulled from Facebook because certain kinds of records – like the state-issued missing person flier – are strong signals that should be verifiable.

Lastly, tensions emerged between the government and the victims, in the form of a lack of transparency on how disappearances were recorded, the limited accessibility to government databases, and the opaque definition of who counts as missing in the official record. These tensions are a reminder of the constraints of the context where we carried out the *data migration,* but also point to the relevance of developing alternative methods to identify cases of missing people. The seven characteristics I developed contribute to establishing an alternative model that details the context of each abduction, thereby shedding some light on how the collective and not only the individual navigate the search for missing people.

Reflections on the Method: Maintaining Context

Data collected from social media platforms need to be understood in its unique political, cultural, and historical context of production and circulation [55, 118]. As previously mentioned, a consistent challenge when using non-video content from social media platforms is preserving meaning through context. The process and findings we present shed light on how qualitative methods can be integrated with computational approaches to maintain the context, to identify potential uses of social media data, and to develop appropriate mechanisms for data migration. This process was deeply informed by previous fieldwork we conducted in both the online communities from which we extracted our data, as well as on the needs of the NGOs that we aim to support [36]. Rather than following a linear progression, the iterative process allowed us to reflect on the insights gathered after each stage, refining the computational methods and contextualizing the resulting data sets in the particulars of the communities that produced those data. This continuous adaptation helped us identify the characteristics in both contexts, online and offline, that enable us to establish correspondence between the work being done on social media in online communities and the data and information needs of NGOs.

Within each key practice – *finding signals*, *scaling signals*, and *verifying signals* – I combined computational and qualitative methods to properly identify and analyze trends and patterns of user-generated content. The former helped tackle challenges of identifying and extracting volumes of data, while the latter methods guided the accounting of issues of representation and interpretation. At the *finding signals* stage, I use qualitative methods to understand the online and offline practices of people and the government when sharing, tracking and addressing cases of abductions. Following an iterative approach at this stage helped to identify not only that people are using Facebook to exchange information about abductions, but they are specifically using the official fliers that the government provides them once a person is reported as missing with the police. Once I understood how people were trading information in these networks, we were able to implement a more suitable method, an image classifier. In contrast, *scaling signals* always required the use of computational methods to gather and process large amounts of data. Lastly, *verifying signals* blended both for a human-in-the-loop computational approach to establish the accuracy of the data analysis for mobilization in the offline context.

In addition to the process developed, the data characteristics I identified contribute to maintaining the context of social media streams during *data migration.* We maintained the locality of data by keeping track of the Facebook communities from which we extracted each post and the details of the location where the person was last seen. This information was identified and recorded through the data characteristic of *situated*. The second characteristic that contributed to maintaining context was the attribute of *descriptive*. This attribute provided insights on the productions of the post and the community where it was circulated and extracted. The *descriptive* characteristic required details embedded within each piece of data such as source, author, and comments and helps establish a connection between online and offline efforts to recover missing individuals. In addition to these characteristics, the attribute of *redundant* helped provide additional accounts of a given abduction, which enabled us to identify whether there were multiple narratives surrounding a missing person. Finally, the attribute of *official*, which refers to the state-issued missing-person fliers that circulated on Facebook, was critical for bridging the online with the offline context because the existence of the fliers meant there was a registered legal complaint. Unlike the other data characteristics, this is the only one that can be considered in itself evidence.

As Loukissas powerfully argues, all data are local, and maintaining this locality demands effort and a kind of care [19]. All together, the data characteristics that we identify in this work aim to maintain this locality and, up to a certain point, encapsulate their context of production and circulation. Our findings have implications for how social media data streams may be migrated and shared across stakeholders, not despite, but because of their heterogeneity.

From Findings To Evidence

The seven data characteristics I identified helped to determine 303 official cases of missing people documented in Facebook group posts. Out of those 303 cases, 276 names were missing from the Mexican national registry of missing persons – the RNPED database.

Establishing this data gap is an outcome of our contribution, the *data migration* process. This process is a methodological pipeline that lets us identify cases of abductions within the social media data stream that were already recognized by the government, and compare them with the official database of missing people. Although the findings establish a data gap, that data gap is not self-explanatory – we do not know what caused or accounts for the substantial discrepancy. It could be an indication of people having been found either dead or alive, or it could be the result of technical or political change [170, 50]. To make these findings useful for NGOs', we need to transition these initial findings into more robust evidence so they can effectively be mobilized with Mexico's legal and policy framework.

For many decades, NGOs have developed methodologies to validate and maintain the credibility of the information they use when monitoring human rights violations. Some of these methodologies involve the use of direct evidence, such as identifying potential witnesses and obtaining representative testimonies, gathering official statements of governments and secondary sources such as local human rights monitors and press sources [171]. These techniques have been changed and adapted based on the context where the fact-findings are taking place, the objectives that sought to be achieved, the target audience, and the source of information [133]. Thus, it takes an accumulation of evidence and techniques, neither a unique approach nor a single piece of evidence is enough to document human rights violations.

When dealing with user-generated content for documenting human rights abuse a significant challenge is to determine what is immediately and long-term relevant and what is misleading or false. These determinations are usually done through human resources coupled with a "deep understanding of the local, social, and political conditions" of the place where the documentation is taking place [56]. Due to this need of human analysis, any evidence or insight obtained from user-generated content is usually considered as the starting point for investigations, rather than seen as an endpoint [55]. Therefore, to fully transition our findings into evidence, would mean partnering with NGOs to identify the kind of arguments they want to focus on and who do they want to persuade, using the data gap we identify as starting point. Some of the arguments we could expect NGOs making with our findings, is asking the government for more transparency on the process of recording cases of missing people. Or NGOs could combine our findings with as many other data sources as possible to identify patterns and trends [55].

Defining what counts as evidence depends on the media by which the data are produced, the characteristics of those data that support re-localizing them in new contexts, the affordances of the platforms from which they come, and the legal and political contexts into which they are applied. Piecing together this entire pipeline is crucial for addressing large-scale social issues, like the human rights crisis in Mexico. As we move toward building tools that support this process we need to integrate the human and computational ways of knowing. Such tools might build on the qualitative characteristics presented here, but they also need to be capable of maintaining the locality of data and should align with local practices.

Challenges and Limitations

Due to the particular conditions in which we conducted our research, it is important to reflect on two main limitations, first the context where the research took place and second the platforms from which we gathered our data. Both the geopolitical context and the platform context shape how our work could be replicated in other parts of the world and with different underlying social media data sources.

Conducting this research in the context of Mexico reveals local implications of using social media as a source to address data gaps. The social and security constraints that residents experience shape to how they adapt the features of social media platforms, depending on their goals and constraints [28]. *Data migration* processes need to account for these particularities as they take different forms in different places. One particular constraint of this context is the fact that many residents in Mexico do not report to the police when their relative is missing due to fear of retaliation and a pervasive lack of trust [160]. Therefore, the challenge is to legitimize and count those cases of missing people that do not have an official complaint, which goes beyond the analysis we completed here but which also may be built on the characteristics we identified.

While Facebook maintains itself as the dominant mediator of a Mexican public sphere, we need to reflect on the implications of focusing our efforts on these types of platforms. This starts with reflecting on the commercial spaces we are supporting, the ecosystems of data that we are reinforcing, and lastly by examining the voices and communities we are missing in these online, public spaces [54]. Tufekci reminds us that although Facebook offers a plethora of information and opinions, this networked public sphere is shaped by the policies, ideologies, legal concerns, and financing models of (largely U.S.) corporate entities [54]. These commercial spaces are subject to a multitude of different legal regimes because they operate in countries with dissimilar and sometimes conflicting notions of free-speech as well as different commitments to liberal discourse. Censorship and moderation practices carefully craft ecosystems of data where only some can freely express themselves. The consequence is that social media platforms have become gatekeepers, defining what constitutes the public and the private sphere, shaping audiences and contributors, and prioritizing or burying content based on a collection of algorithmic outcomes that at illegible and inaccessible to individuals using these platforms [54, 172, 173].

Additional challenges include overcoming the reproduction of inequality and the exclusion of those most affected by a crisis by naively turning to social media data streams [174]. The design of any automated system to facilitate the migration of data from an online context to an offline context – and specifically when seeking to establish robust evidence for use in legal or policy interventions – requires us to consider alternatives that avoid disrupting local practices and creating new dependencies. For example, preventing disruptive local practices would imply the design of tools capable of grappling with the range of conceptions and uses of data of those who capture, analyze, and draw conclusions from them.

In this light, the use of Facebook posts as the primary source of data becomes an inherent issue of dependency. Therefore, to avoid this type of dependencies, we will need to design mechanisms to store and maintain the Facebook posts, holding their contextual origin such as comments and names of the Facebook group where they were created.

5.4.1 Conclusion

Instrumentalizing social media data as evidence means understanding the complex ecosystem of actors and institutions intertwined with the political and historical context of databases and infrastructures that define what counts as evidence. In this study, governments' databases and norms guided the definition of evidence when examining the data from the Facebook groups. I determined the type of posts on Facebook that could become evidence of missing people based on how these incidents are being measured and classified by the government institutions. Similarly, I established the data gap on the number of missing people by using the government databases as references.

However, to make social media data actionable, it is imperative to partner with organizations and other stakeholders who could benefit from the findings and insights those data may provide. To understand their standards of evidence and work practices, identify the kind of arguments they want to focus on and who they want to persuade. In this sense, it can be said that the value of social media data as evidence is determined by what organizations need.

In the next chapter, I present a study conducted in partnership with a civil organization where we examined data from Twitter to characterize how citizens, local governments, and grassroots in Mexico City collaborated to address the social and economic impact of COVID-19. With this study, I began to tease apart the implications of integrating the methodology developed in this chapter into the pipeline of organizations interested in using social media data as evidence.

CHAPTER 6

STUDY THREE: HUMAN DATA WORK

The initial motivation of this study was to examine what would entail following the methodology previously developed within the context of an organization. As I conducted the study, it became evident that examining data construction and human interpretation were essential to surface the possibilities and limitations of the knowledge we can extract from social media data. In this study, I investigated the role of human interventions and interpretations in mobilizing social media data from their platform of origin to the institutional context of a civil organization. During seven months of remote fieldwork with the Accelerator Lab, we followed a mixed-methods approach to examine data from Twitter to identify grassroots initiatives addressing the social challenges of the COVID-19. The findings of this research illustrate the role of human interpretation and organizational practices in defining, negotiating, and interpreting the meaning of social media data in cooperation with my partners to obtain insights. The contribution of this study is to show that the meaning of social media data is not defined in advance; instead, it is contingent and negotiable on the practices and needs of the organization that might benefit from the analysis of such data.

6.1 Introduction

The increase in using digital traces for data analytics and human behavior prediction has risen in part due to the widespread public availability of large datasets of user-generated content and because social media platforms have become places of convergence. However, disciplines such as Science and Technology Studies (STS) and Critical Data Studies have surfaced various limitations and restrictions when using user-generated content to examine human networks and communities [16, 175, 176, 177, 178].

Scholars from these disciplines urge us to recognize the role of human intervention in the construction and interpretation of data, foregrounding the technical and subjective constraints of data analysis divorced from data collection. These observations become more prominent when considering the unique challenges that the social, cultural, and temporal context of social media data poses. As I mentioned before, capturing and documenting data context is imperative: it is impossible to extract social media data from its context of production and generate analysis that makes any claim of objectivity or accuracy [16].

The research I present in this chapter is concerned with two aspects that have received little attention when making sense of human behavior via digital traces. First, we are interested in examining the role of human interpretation when mobilizing social media data from their platform of origin to the site of use. The decisions of which attributes of social media count, which are ignored, the association of those attributes with human behavior, and the definitions of evidence are all subjective choices bounded to disciplinary practices. Therefore, in this research, I examine how human intervention and interpretation are inscribed in the process of mobilizing social media data from their platform to the institutional context of an international NGO.

Second, I am interested in understanding how the organizational practices and needs of stakeholders, who might benefit from the insights of social media data, influence the selection and interpretation of that data. Building on previous observations of STS scholars, I argue that mobilizing user-generated content to an institutional context entails transforming the content's legibility. As data are mobilized between sites of practice, the socio-cultural values that practitioners follow and use to justify their data practices ultimately affect the meaning ascribed to the data.

In order to illustrate how these cross-context analyses take shape, I report on seven months of remote fieldwork and partnership with an international NGO based in Mexico City. With my partners, I followed a mixed-methods approach that examined Twitter data to identify grassroots initiatives addressing the social and economic challenges of the COVID-19 health crises. This research offers a critical account of the collaborative work practice that entails mobilizing social media data from their place of production to the institutional context of my partners so that they could be used as evidenced to inform their work. Our findings illustrate the role of human interpretation and organizational influence in the process of *defining, negotiating, and interpreting* social media data. By offering this account, we aim to provide evidence on how data construction and interpretation are essential to surface the possibilities and limitations of the knowledge we can extract from social media data.

Our research is in conversation with emerging perspectives in the field of HCI and CSCW that are concerned with the disciplinary interpretations that arise between data practices and data subjectivities [25, 24, 179, 23, 180]. This recent work understands data as a social construct—rather than a neutral entity—from which interpretation heavily depends on the tools and disciplinary data practices [25, 30, 24]. As a result of these interpretive disciplinary gaps, researchers have established the need for developing practices and methodological approaches capable of surfacing [119], capturing [23], and recognizing the human interventions and social constructs of data [25, 24]. This work has helped bring to light how interventions and constructs move across sites of practice and infrastructures [121]. These critical examinations mostly have taken place in data science; however, little has been done to examine the role of human interpretation and subjectivity in *mobilizing* social media data from digital platforms to institutional contexts such as NGOs. In this research, we surface some parallels between previous work on data science and social media data, specifically to inform the work of NGOs.

Lastly, mobilizing social media data into the institutional context of our partners was a collaborative effort guided by their goals and epistemologies, and constrained by the inherent limitations of user-generated content. Our contribution is to show that the meaning of social media data is not defined in advance; instead, it is contingent and negotiable on the practices and needs of the organization that might benefit from the analysis of such data.

6.2 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 [181]. 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 [182, 183]. 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 [184]. At the time of conducting this research, 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 [185].

While the pandemic was prompting official response across all level of government in Mexico, we began seven months of fieldwork and collaboration with *an international organization (anonymized for review).*[1] This organization focuses on working closely with local stakeholders to identify community-level solutions that have the potential to accelerate development. During the pandemic, identifying successful local initiatives that might be scaled or translated to other locations was a priority. To do this, the organization started exploring new methods to identify innovative sources of information that could facilitate high-level decision-making on complex problems.

In May of 2020, the organization launched the *COVID-19 Social Inventory Project* as part of their overall pandemic response. The project consisted of exploring social ties by identifying citizen initiatives that originated in response to the effects of COVID-19 across communities in Mexico City. The initiatives included any individual or collective action responding to the COVID-19 health crisis. As a guiding framework, the *organization* was interested in identifying three different types of social capital: bonding, which connects people within a community; bridging, which connects across communities; linking, which

[1] From here on we refer to our partner organization as the *organization* or *our partners.*

connects communities with government. Additionally, the *organization* focused on identifying the similarities and differences between the sixteen *alcaldias* that make up Mexico City [2].

To anchor the work on the ground in Mexico City, the organization partnered with the Youth Institute of Mexico (INJUVE) [3] and recruited members of the institute so they could help distribute the survey. By partnering with INJUVE, the *organization* sought to understand how to collaborate with and better understand the perspective of the young population in Mexico City, as well as to identify grassroots initiatives addressing COVID-19.

6.3 Methods

Our partners' central study consisted of a survey to identify citizen initiatives and examining various official datasets to measure social capital. In parallel, and based on our expertise, our partners at the organization tasked us with using social media as a complementary and alternative source to identify initiatives and provide insights on the various types of social capital across Mexico City.

6.3.1 Partner Data Collection

The survey designed by our partners consisted of seventeen questions, all of them were multiple choice with three open-ended questions. The design reflected the details of the initiatives that were important to the organization. We used the survey to inform our social media data collection process.

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

[3] In Spanish: Instituto de la Juventud de la Ciudad de México (INJUVE). INJUVE was established in 1999 with the purpose of coordinating public policy aimed at young people in Mexico City. Site: https://www.injuve.cdmx.gob.mx/

Table 6.1: 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.

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.*

First, we categorized the survey questions based on the information our partner sought to capture. As a result of this categorization, we identified six characteristics of the initiatives that were relevant for the *organization* including 1) purpose of the initiative, 2) target population, 3) location, 4) identity of the organizers, 5) time, and 6) contact information for the initiative. These six features were relevant to the *organization*'s goal to track initiatives and guide future community interventions. Later in our analysis, we captured these features from initiatives identified from social media. We discuss this in more detail in the Findings section.

To understand what counted as a citizen initiative, we analyzed the responses to one of the open-ended questions from the survey, which asked for a description of the citizens' initiatives. We read through 147 responses, which offered rich narratives of the initiative's objectives and target population. Based on this analysis, we concluded that an initiative referred to individual or collective actions that aimed to address or reduce the negative impact of social, economic, or health problems caused by the COVID-19. Additionally, we compiled a list of hashtags, keywords, and social media accounts of organizers of initiatives coordinated via social media. We organized this information based on the categories of the initiatives' purpose, which included food, health, education, labor, and public communication. Table 6.1 shows a sample of the attributes we collected for initiatives that focused on food.

6.3.2 Data Collection

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 in the platform. Therefore, we decided to use Twitter because it allows for data collection and, after Facebook and WhatsApp, it is the most used social network in Mexico. According to recent statistics, there are eleven million Twitter users in Mexico, representing 60% of Internet users between 16 and 64 years old [186].

Using the Twitter API, we collected data between February 28th and May 17th,2020 covering the two initial phases of COVID-19 in Mexico City. According to the health ministry in Mexico, the first epidemiological phase, which began on February 28th,2020 consisted of identifying imported cases. The second epidemiological phase, which was declared on March 24th and lasted until May 17th, was marked by community spread. We filtered our collection of tweets using the name of the sixteen *alcaldias* of Mexico City in combination with the hashtags and keywords collected in the preliminary analysis. In total, we collected 300,361 tweets.

We acknowledge that all social media dataset are partial, and collecting data from Twitter constrains the perspectives, experiences, and insights that can be obtained for analysis [16]. Despite these limitations, we opted for Twitter due to its widespread adoption among citizens, making it a powerful tool to identify instances of civic responses and initiatives that contributed to the city's recovery process.

6.3.3 Natural Language Processing Tool

We implemented a natural language processing (NPL) tool for Spanish language to support the analysis of the collected data from Twitter. The tool was based on word embeddings using the *word2vec* algorithm [187, 188]. During word2vec's training phase, a large corpus of text is mapped to a vector space with each unique word being assigned a *word vector*. Semantic meaning of the text corpus is preserved by placing word vectors close to one another in the vector space when they share common contexts [187]. Our tool made use of the Spanish Billion Word Corpus and Embeddings (SBWCE) pre-trained model, which comprises nearly 1.5 billion words [189].

The input for our tool was a set of cleaned tweets from an *alcaldia* and a set of sentences describing a given initiative. Cleaned tweets and initiative sentences were mapped to vectors using word2vec with SBWCE. Then, each tweet vector was tested for cosin similarity with each initiative's sentence vector. In this context, a cosine similarity close to one indicates a tweet was semantically similar to a given initiative's description sentence. This tool was used to reduce the overhead of manually analysing this relationship.

6.3.4 Data Sampling and Analysis

We conducted the data analysis in three stages. First, we analyzed the dataset of tweets of each alcaldia using the word2vec tool. We obtained a single output file per alcaldia with approximately 1000 and 1500 tweets with content related to initiatives. Using NLP helped us to narrow the number of tweets we analyzed in the following stages. The second stage consisted of further reducing the number of tweets for qualitative analysis. Using the randomize tool in Excel, we selected a sample of 100 tweets per alcaldia. Then, following our partners' definitions of bonding, bridging, and linking, we categorized each tweet into one type of social capital. After the first round of analysis, we continued sampling tweets in batches of 100 until we reach saturation of each *alcaldia*. On average, we analyzed between 300 and 350 tweets per *alcaldia*, more than 4,800 tweets in total. In section 5.1, we describe how we associated tweets with social capital types.

As we associated the tweets, we gathered additional information that we used later to guide our interpretation. We documented the supplementary information in a log file, which is described in Table 6.2. In section 5.2, we provide a detailed description of our rationale to collect and document additional information.

In the last stage, we used content analysis to identify commonalities, distinctions, and relationships amongst the local responses that citizens, governments, and grassroots efforts were coordinating per *alcaldia* to address the challenges of the COVID-19 crisis. Our emerging themes reflected the different manifestations of linking and bridging across alcaldias.

6.4 Findings

The findings we present surface the various decisions we made with our collaborators to effectively mobilize the data from Twitter into the institutional context of our partner organization. Our observations highlight how using computational tools to extract and make sense of social media data is not purely the result of the technical application of machine learning and automation, but involves countless trade-offs, decisions, and assumptions that conceal a set of values and interests of those doing the work of categorizing and interpreting the data. While these decisions and activities are rarely discussed when using digital traces as a source for characterizing and predicting human behavior, they heavily influence our understanding of what social media data can tell us. Our findings provide evidence on how those decisions are bounded by the organizational needs and practices of the institutions guiding the work of data, challenging the notion of objectivity when using computational techniques to extract knowledge from social media platforms.

6.4.1 Calibrating Content Association

As described above, our partner's goal was to understand how people were organizing to respond to the effects of COVID-19 and to identify which *alcaldias* had a higher social capacity for collaboration and grassroots innovation. To inform their work, the organization used the framework of social capital that distinguished between three types, including

Table 6.2: Categories and subcategories of the data from Twitter that were collected and documented in a log file

Category	Subcategory
Tweet URL	Address of the tweet
Date	Day, month and year when the tweet was published
Social Capital Type	Linking or Bridging
Linking Subcategories	Food, Health, Public spaces, Informal commerce, Formal commerce, Education, Information, Economic assistance, Public transportation, Public services
Bridging Subcategories	Food, Health, Donations, Local commerce
Reaction	Citizen's approval, Citizen's complaint, Government response or communication
Publisher	Government, Citizen, Public figure, Newscast /Newspaper/Journalist, Non-profit organizations, Enterprise, Educational institution
Target Population	Elderly people, Health personnel,Indigenous groups, Citizens, Marginalized or homeless populations, Merchants, Farmers, Children, Youth, Women, LGBTQ, Students
Organizer	Government, Citizens,NGOs, Enterprise, Educational institution
Place	Alcaldia's name, Neighborhood
Text	Content of the tweet

bonding, bridging, and linking. While we had an explicit definition for each type of social capital, the representation of these phenomena in Twitter data was diverse, requiring us to define what counted as a canonical example of these behaviors, and then analyze and interpret each tweet under a common understanding with our partners.

As we collected data from Twitter, we showed our partners preliminary findings to agree on the interpretation and categorization. Together, we decided which type of tweets encoded the types of social capital our partners were interested in identifying to inform their future interventions. Thus, *calibrating interpretation* consisted of associating online evidence with our partner's definition of bonding, bridging, and linking, and then analyzing the context surrounding each of the tweets to gain deeper insight into their meaning.

Associating Twitter Content with Types of Social Capital

We held a calibration session with our partners to identify canonical examples of tweets representing the three types of social capital. In the calibration session, we showed our partners various tweets that exemplified what we considered to fall into the category of linking, bridging, and bonding. Initially, we focused on defining what kind of content [4] we counted as evidence of each type of social capital.

Our partners characterized the social capital of linking as *the formal and informal relationship between citizens and government authorities*. Based on this definition, we initially interpreted linking as programs and actions organized by government staff to serve the citizens and to reduce the impact of COVID-19 (see figure Figure 6.1). [5] Thus, in the calibration session, we showed our partners tweets with content that reflected such activities. Some examples included: government programs advertised on the mayors' accounts of the alcaldias, newspapers reports, and citizens' accounts on government aid programs and intermittent activities carried out by the government. Our partners agreed that these types of tweets should be classified as linking because their content described actions organized by the government: going forward, we classified tweets that reported on government programs and informal activities as linking initiatives. Figure Figure 6.1 and Figure 6.2 show examples of the tweets we categorized as linking.

Communication between local government and citizens varied greatly among *alcaldias*. As we progressed with our analysis, we identified tweets that reflected unusual interactions between government and citizens. We found that out of the sixteen *alcaldias*, six of them use social media platforms to communicate with their constituents. For example, on Twitter, the mayor of the alcaldia Xochimilco advertised the creation of WhatsApp groups to communicate directly with residents. Another example was the mayor of alcaldia Azcapotzalco, who broadcasted YouTube videos and advertised them on Twitter, providing updates on government programs. We also found tweets that showed communication between citizens and police (see figure Figure 6.2).

We shared these tweets with our partners to decide together whether or not we should

[4]"Content" refers to the text, images, videos, or links included in the tweet text.

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

classify them as linking, because the content of some of the tweets was unrelated to COVID-19. For example, the tweet in figure Figure 6.2c shows a conversation that unfolded between the police and a citizen who reported people drinking alcohol on the streets. We considered these interactions unconventional because historically in Mexico citizens do not report crimes to the authorities. For the last ten years, the results of the National Survey of Victimization and Perception of Crime [6] have reported 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. Therefore, the engagement we observed on Twitter between civilians and government staff of alcaldias was unexpected and we consider it worth being recorded. Our partners agreed on the importance of capturing these interactions, so we broadened the criteria for classifying tweets as linking and included the tweets that reflected interactions between government and citizens, even if the conversations were not directly related to COVID-19.

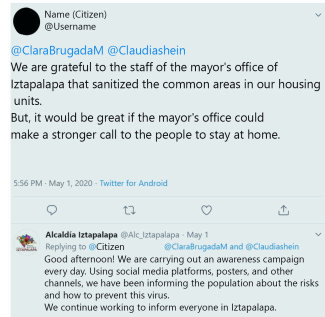
As a result of our analysis, we identified ten different subcategories of initiatives, aid programs, and interactions between government and citizens. These categories include 1) food, 2) health, 3) public spaces, 4) informal commerce, 5) formal commerce, 6) education, 7) information, 8) economic assistance, 9) public transportation, and 10) public services.

After defining the criteria for linking, we worked with our partners to define the criteria for bridging social capital, which our partners understood as *the relationship between groups of people who seem to have less in common*. For the calibration of bridging, we showed our partners the tweets in figure Figure 6.3 because we considered them to reflect a connection between different groups. However, our partners decided that only tweets similar to figure Figure 6.3a and figure Figure 6.3b should be classified as bridging, because those tweets made it possible to identify and differentiate the group of people providing

[6] In Spanish Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE).

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

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



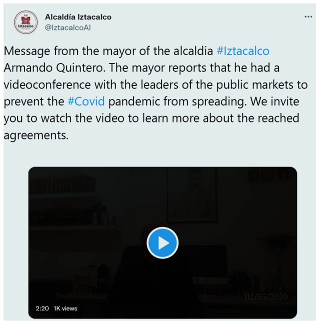
(c) Tweet from a citizen thanking the mayor's office for the sanitizing squads

Figure 6.1: Tweets showing targeted actions that the mayor's office of the *alcaldia* Iztapalapa undertook regarding food and public health.

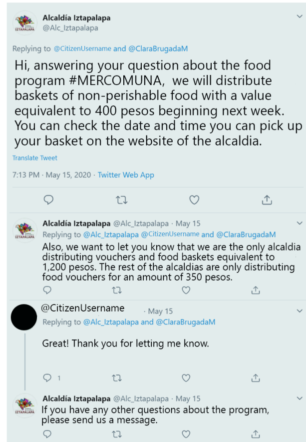
help and the group receiving aid. Adding further nuance, our partners did not consider the tweet in figure Figure 6.3c an appropriate example of bridging because it was a business focused on payments, package delivery, and food collection rather than voluntary assistance. As a result of our analysis we identified four subcategories of bridging including: 1) food, 2) health, 3) donations, and 4) local commerce.

Thus, we established that to classify tweets as bridging, their content would have to provide us with enough information to 1) identify the group or individual providing help and as well as the one receiving help and 2) determine that it is a voluntary donation and not a business. In section 5.2, we describe in more detail how we determined those factors.

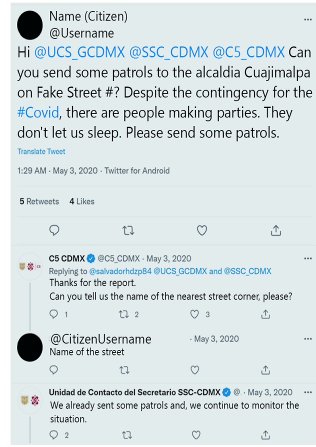
Lastly, we discussed with our partners how to establish the criteria for bonding. Our partners defined bonding social capital as *the connection among individuals that share similar characteristics*. In the calibration session, we showed our partners the tweet in figure Figure 6.4 describing how relatives of inmates organized to donate medicine to the prisons. Since the family members organized the donations, we inferred that aid was occurring between similar groups. However, we found it difficult to identify more tweets in which the content reflected an interaction between equal groups (e.g., relatives or close friends). Therefore, we decided with our partners not to continue searching for this type of social



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



(b) Conversation between a citizen and Twitter's official account of *alcaldia* 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 6.2: Examples of multiple engagement strategies between government staff and citizens.

capital, because the information contained in the tweets was not enough to conclude that similar groups of people were helping each other.

6.4.2 Documenting and Interpreting Additional Content

During the calibration sessions, we also decided with our collaborators on the additional information we needed to collect to interpret the content of the collected tweets. As part of our analysis, we maintained a log file where we recorded and categorized the details of each tweet, which later guided our interpretation and association of data with the social capital of linking and bridging. Table 6.2 shows the Twitter attributes and the content categories we recorded for each tweet. This content helped us understand the context of the production of the tweets, which in turn influenced our interpretation of the data.

We categorized the supplementary information we considered in our analysis into four groups: 1)publishers, 2)reactions, 3)organizers, 4)target population.

(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 6.3: Examples of the type of tweets we associated with the category of bridging

Figure 6.4: Example of a tweet categorized as bonding

Publishers and Reactions

Our partners pointed out that as part of the interpretation and categorization of data we needed to consider the publisher of the tweet. Specifically, when defining what counted as an example of *linking*, our partners raised the issue that tweets posted by citizens cannot be interpreted in the same way as tweets posted by government personnel.

For instance, the tweets in figure Figure 6.1 classify as linking because they describe government initiatives. However, each tweet was published by different actors. The tweet in figure Figure 6.1a was published by the mayor of the *alcaldia* Iztapalapa, the tweet in figure Figure 6.1b was published by the Mexican newspaper La Jornada, and the tweet in figure Figure 6.1c was published by a citizen.

Our partners argued that tweets describing government initiatives posted by the mayor's office or government Twitter accounts could be part of their marketing campaign rather than a genuine interest in helping citizens. In contrast, they considered tweets posted by newspapers or citizens describing government actions as more meaningful to understand the impact of the government efforts because they might reflect people's perceptions of the benefits of the initiatives. Therefore, we decided to document and categorize the type of account publishing the collected tweets. As our analysis evolved, seven emerging categories of publishers emerged reflecting the different types of accounts sharing initiatives. These categories include 1) government, 2) citizen, 3) public figure, 4) newspaper, 5) NGOs, 6) enterprise, and 7) educational institutions.

Additionally, during the calibration process, our partners pointed out the relevance of considering people's reactions to the tweets, specifically the replies to the tweets containing information on government initiatives addressing COVID-19. Thus, we documented citizens' replies into two categories:*citizens' complaints* and *citizens' approval*. Examining these reactions gave us some insight into citizens' perceptions of government performance and gave us a more nuanced understanding of the relationship between government and residents.

Organizers and Target Populations

Identifying both the organizers and the target population of the initiatives we found on Twitter were two crucial attributes that defined whether we classify a tweet as evidence of linking or bridging.

Depending on the content of the tweet, the categories of organizers and targeted populations had to meet different criteria. For us to classify a tweet as an instance of *bridging*, we ensured that the actors coordinating the initiatives belonged to a different group of people from the intended target population. In contrast, we categorized as *linking* those tweets where we identified cues that indicated the initiative organizer was the local government. Through our analysis, we identified five different types of organizers and twelve different categories of target population. Table 6.2 shows these categories.

As we examined our data, we found that most of the collective responses were organized by non-profit organizations through the coordination of donations. And, to a lesser extent, we found examples of citizens organizing collectively on their own. Based on our analysis, we concluded that the collective responses on behalf of non-profit organizations are a sign of capacity for recovery that is coming from outside rather than within the community.

For the tweets we classified as *linking*, identifying different targeted populations helped us understand common pressing concerns shared across local governments. For example, the governments of the *alcaldias* Benito Juarez and Azcapotzalco implemented initiatives and distributed funds to help elderly people, and women, respectively. Among the sixteen *alcaldias* in Mexico City, these two were the only ones that developed tailored programs to help these populations. Learning about these programs gave us some insight into the priorities of the local governments of both alcaldias and how those are translated into recognizing that the pandemic affects vulnerable groups differently.

6.4.3 Situating Social Media Data

The last decisions we made consisted of defining how to organize the outcomes of the two previous analyses to inform our partners' interventions. We recognize that data from social media platforms are always incomplete. The observations we collected from Twitter will never be exhaustive of the citizens and government initiatives to address the COVID-19 crisis in Mexico City. However, they provided evidence of collective responses, assets, and particularities of each alcaldia that otherwise would have been difficult to identify. Considering such limitations and constraints of using Twitter data, we decided that the most appropriate approach to make sense of the data collected was to provide a holistic

Table 6.3: Categories of insights

Category	Description
Social Characteristics	**Local Problems:** Tweets that refer to any social problem in the alcaldia. **Informal Economy:** Any reference to street vendors or informal commerce in the alcaldia.
Linking	**Population:** Population to which the alcaldia's government focuses more by offering them special aid. **Relationship between citizens and government:** Tweets that reflect the relationship between citizens and the mayor's office across alcaldias. **Government aid programs:** Tweets that describe or promote aid programs lead by alcaldia's local government.
Bridging	**Collective organization and response of citizens:** Tweets that describe or promote citizen's initiatives to address the social and economic challenges of the COVID-19 health crises. **Relationship with other municipalities:** Tweets describing instances of any type of collaboration between alcaldias. .

analysis of each alcaldia.

Our final analysis consisted of an individual qualitative overview of each alcaldia and a comparison of its strategies and priorities. We characterized each alcaldia by summarizing individual observations on linking, bridging, and social characteristics.

Table 6.3 describes the elements we considered for each category. The characterization of linking consisted of a description of local government strategies when responding to populations' needs, the populations that were a priority for each alcaldia, and the communication between citizens and government. Bridging consisted of the collective efforts of communities who organized and leveraged their resources to address the social and economic consequences of the pandemic. Lastly, social characteristics consisted of describing local problems rooted in the alcaldias. In this section, we provide a short overview of the highlights of bridging and social characteristics that we communicated to our partners.

Bridging

Xochimilco is one of the few areas in Mexico City where farmers harvest vegetables using a traditional technique named chinampas: an agroecosystem of pre-Hispanic origin artificially built in areas of the lake of Xochimilco [190]. Residents in Xochimilco from this alcaldia organized to sell and distribute their products, establishing commercial networks across Mexico City to create impact beyond their immediate community. As we analyzed the tweets in our data set, we learned that the restaurants in Mexico City in the past were the primary consumers of the produce from Xochimilco. However, due to the pandemic, most of the restaurants in Mexico City closed, impacting the farmers. In response to the economic fallout, farmers in Xochimilco rapidly organized during the first stages of the pandemic to implement the sale and shipping of vegetables across Mexico City, creating supply networks with other alcaldias. We identified this population through multiple streams within our tweets: some were published by the farmers themselves as they organized to distribute their products through online channels, while others came from the local governments as they encouraged people to consume the local produce. The organization of the chinamperos illustrates how different networks converged to respond to the health crisis. Learning about the strengths and capacities of the populations helped our partners to identify the levels of collaborative capacity of communities in Mexico City.

Social Characteristics

In ten out of the sixteen alcaldias, we found an overwhelming number of tweets that referred to water scarcity. Through our analysis, we learned that for many alcaldias the shortage of water had been a persistent problem for the last ten years, and it became more acute with the pandemic. We found evidence that the residents of the alcaldias Tlahuac and Cuajimalpa organized to address water scarcity by collectively buying water tanker trucks. These observations reflect the local challenges of alcaldias, as well as the community resilience of their inhabitants.

Another local challenge we identified was the shortage of medical supplies. This problem was most prominent in the tweets of the alcaldias with the majority of hospitals in Mexico City, alcaldias Tlalpan and Gustavo A. Madero. As a result of the shortage, the medical staff of the hospitals frequently organized public demonstrations that involved closing avenues and streets from the surroundings of the hospitals. Through the tweets, we learned about the demands of the medical staff as well as the impact of the protests on the neighbors, patients of the hospitals, and the city in general.

6.5 Implications: Framing Knowledge through Organizational Practices

Recent research on CSCW and HCI has begun to examine how collaboration intertwines with technical forms of work across the multiple practices in data science. As part of this nascent inquiry, researchers have shown that the definition and construction of ground truth is a human endeavor [191, 192] and a negotiation [26] constrained by organizational practices, business goals, and existing technologies [193, 194, 23]. In this research, we defined and calibrated with our partners what counted as ground truth while carefully considering the limitations of representation and context when interpreting social media data. In addition to the close collaboration, the unique organizational practices of our partners determined what counted as valid and reliable knowledge, which constitutes their *culture of objectivity:* the process of constructing knowledge that is internal to institutions and disciplines [195].

The organizational practices and epistemologies of objectivity of those doing the work of data determine what counts as evidence and ground truth, which in turn affect the mobilization of data. Previous work leveraging social media data has primarily focused on developing and perfecting computational methods to characterize human behavior without considering the cultures of objectivity of those institutions that might benefit from such research in an attempt to see without inference and interpretation. When mobilizing social media data into the institutional context of organizations, it is imperative to consider their culture of objectivity that prevails and its influence on the definition of ground truth and evidence.

My partners embraced multifaceted ways of knowing beyond statistical and quantitative reasoning, which translated into establishing partnerships with multiple entities and organizations to develop new methodologies that allow them to address complex development problems. My collaborators showed flexibility and willingness to discuss and examine the evidence of social media, recognizing the challenges of classifying these data into discrete, numerically determinable categories. Therefore, rather than obtaining a quantifiable overview of citizens' initiatives, social media data was seen as a tool to access a more descriptive narrative of collective organization across Mexico City. The insights obtained from social media platforms revealed strengths in communities that were not visible in the more traditional sources that my partners used in parallel, such as government statistics.

6.6 Implications: Tools For Documenting the Human Data Work

The decisions we made to mobilize social media data were constrained by the characteristics of Twitter and by my partners' interests. We classified and interpreted data based on patterns of inclusion and exclusion defined by my partners' needs. The outcomes of these choices impacted the claims associated with the data analysis. Developing an accountable understanding of the decisions involved in mobilizing social media data into the contexts of organizations is necessary to promote a thoughtful examination of the situated data practices that are involved.

In conducting this research, documentation was instrumental in the process we followed when making sense of data. The documentation we kept served as a trace of the human work on data making and interpretation, which translated into transparency and accountability when extracting insights from social media to inform other actors and stakeholders about our findings. Moreover, we centered our documentation efforts on recording decisions about what counted as evidence of the types of social capital from the calibration sessions. To this end, we used spreadsheets linked to memos to record the classification of tweets and to document the process of calibration and interpretation with my partners. However, these tools made collaboration difficult and were not sustainable for future iterations.

To move this research forward, I proposed designing tools that facilitate connecting how the decisions for data collection, definition, and interpretation respond to the organizations' goals doing the work of data. I envision designing templates of worksheets that encourage practitioners to reflect on the type of data or evidence that interests them the most (e.g., personal narratives or statistics), the goals of their project, and their perceived constraints of the social media platform. Then, practitioners could describe how these elements influence each other at three stages of mobilizing data —when defining the social media platform for data collection, defining what counts as evidence, and defining how to interpret the collected data. These tools would support the documentation of how these elements are intertwined. Providing evidence of these connections will allow us to identify the precise moments that human interventions are happening and their rationale.

CHAPTER 7
DESIGN OF THE TOOLKIT

For decades the development of toolkits have play an important role in the field of HCI as *"generative platforms"* that enable creating different solutions by *"reusing, combining, and adapting"* the components or structures provided by the toolkits [196, 197]. Moreover, within HCI, toolkits function as artifacts that translates conceptualizations and provide supporting infrastructure to test technical concepts and paradigms [198].

In the context of non-profit organizations and the development sector, there has been a proliferation of toolkits to increase resilience in the communities where they operate [199], improve community work [199], facilitate design thinking methodologies, support social innovation [200], monitor and evaluate their social impact [201], promote and protect human rights [202] and support activists against the most common types of digital emergencies [203], just to mention a few. Regardless of their topic or context, at their core, toolkits are a chosen collection of resources, materials and tools designed to formalize a particular process to serve a specific purposes [204, 205].

In this chapter, I describe the design of the toolkit *Bitácora* [1] addressed to practitioners working in civil and non-profit organizations interested in harnessing the potential of data from Twitter to identify local capacities, monitor community crises, and develop interventions based on local practices. The purpose of the toolkit is to guide practitioners on distinguishing community perspectives and the context of data production when collecting and interpreting social media data, specifically from Twitter.

[1] The word *Bitácora* means in Spanish log book.

7.1 Why a toolkit?

The findings of my completed research show that communities increasingly use social media platforms to exchange experiences and organize to address local crises. The data gathered in these platforms offer valuable situated knowledge to NGOs working on development projects that would be difficult to capture otherwise. However, as I have already described mobilizing social media data into the institutional contexts poses challenges associated with the nature of the social media data —issues of representation and accuracy— the definition of evidence in a determined context, and the challenges that entail reusing the data including maintaining context and human intervention [28, 29].

To address the challenges that the mobilization of social media data entails, I synthesized the learnings gathered in my previous research and translated them into a set of instruments and practices formalized into a toolkit that will support practitioners in integrating social media data into their work. Toolkits are not just an assortment of tools and materials. As Mattern argued, kits are designed to make an argument for best practices, modeling a particular vision of the world and providing us with the tools to interfere in that world. With this toolkit I call practitioners for a critical and situated approach to analyze social media data.

7.1.1 Bitácora

The toolkit consists of computational tools that enable searching, collecting, and analyzing data from Twitter. Additionally, I designed a manual that encourage practitioners to reflect on the definitions of evidence in their work, instruct them on the trade-offs of data collection and provide the tools for users to document and reflect on their decisions when associating social media data with a specific phenomenon. The toolkit includes various worksheets that allow practitioners to document the decisions and trade-offs when mobilizing social media data into institutional contexts.

7.2 Components of the Toolkit

The toolkit consists of a platform that includes two computational tools to search, download and analyze Twitter data. Additionally, the toolkit includes a manual that guides readers on using the platform's tools to search, collect, interpret, and integrate data from social media platforms to inform their work with a critical and contextualized perspective.

7.2.1 Platform

The platform is a website (http://bitacora-tk.desarrollo.in/) that hosts the computational tools, which I refer to as Module 1 and Module 2. The former consists of a user interface that allows users to search and download Twitter data, and the latter enables the analysis of tweets at scale using Natural Language Processing (NLP). In combination, both modules aim to ease the process of filtering tweets to identify content similar to the problem users are examining.

Implementation

The toolkit is implemented in the Python programming language version 3. The user interface is developed with the Holoviz's Panel library [206] which provides a rich set of graphical widgets and methods to manage the files produced by the toolkit and functionally deploy the resultant code as a webpage.

Implementation of the Module 1

The functionality of Module 1 is to access, retrieve and save Twitter data from user defined queries. Access and search methods are based on the Twitter API v2 [207]. Twitter API credentials allow users to retrieve and analyze Twitter data and offer different types of access for developers and academic researchers. The toolkit is implemented using the academic research credentials, which allow retrieving up to 10 million tweets per month and 100 requests per 15 minutes.

Module 1 was implemented with the Tweepy library which provides a high-level abstraction of the Twitter API v2 thus reducing development time [208]. Additional functionality is implemented to translate the input of the Panel widgets into Twitter queries syntax following the Twitter API v2, and to display and save the results of the query in a Panel's tabular form and csv files [209].

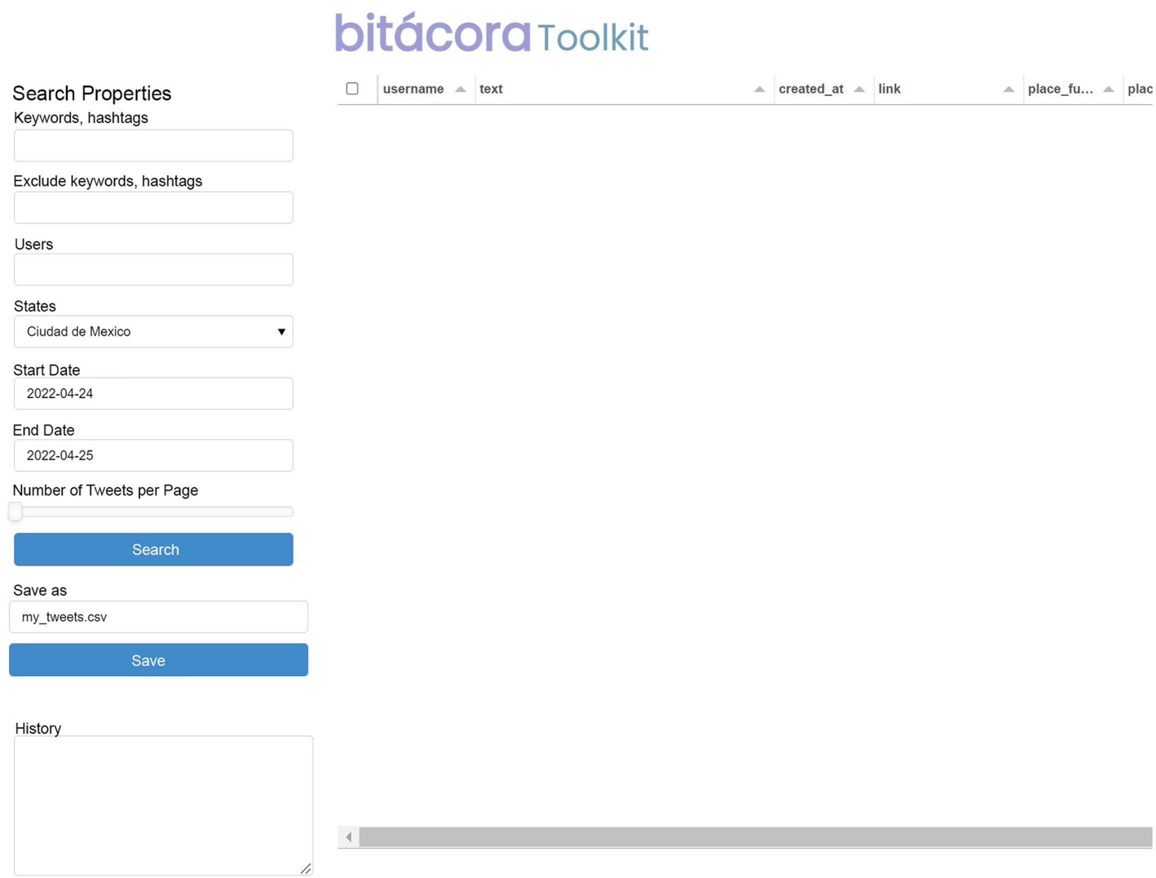


Figure 7.1: User Interface of Module 1

Implementation of the Module 2

The functionality of Module 2 is to support the analysis of Twitter data collected with the Module 1. We implemented NLP for Spanish Language based on word embeddings using the word2vec algorithm [187, 188] with the Spanish Billion Word Corpus and Embeddings (SBWCE) pre-trained model [210], and the Gensim library [211]. The input of Module 2 is two files in csv format. The first file contains the Twitter data retrieved with the Module 1, and the second includes sentences related to the problem being investigated. Both files are cleaned and tokenized before being loaded into memory as dataframes. Then, we calculate the vector representation of each tokenized text using the SBWCE pre-trained model. At this point, Module 2 is ready to evaluate the similarity between the tweets and the problem's sentences. For simplicity, we implemented the `dot product` as a similarity metric. In this way, when the result is close to 1 the tweet and a given problem's sentence are semantically similar, likewise, if the result is close to zero, they are semantically dissimilar. We implemented different Panel widgets to allow the user to control similarity calculations and display the results in a Panel tabular form. [2] Finally to provide and manage users access to the toolkit, the solution was deployed in a virtual machine using the Digital Ocean's cloud services. The ultimate purpose of the Module 2 was to narrow the number of tweets that practitioners will analyze in following stages.

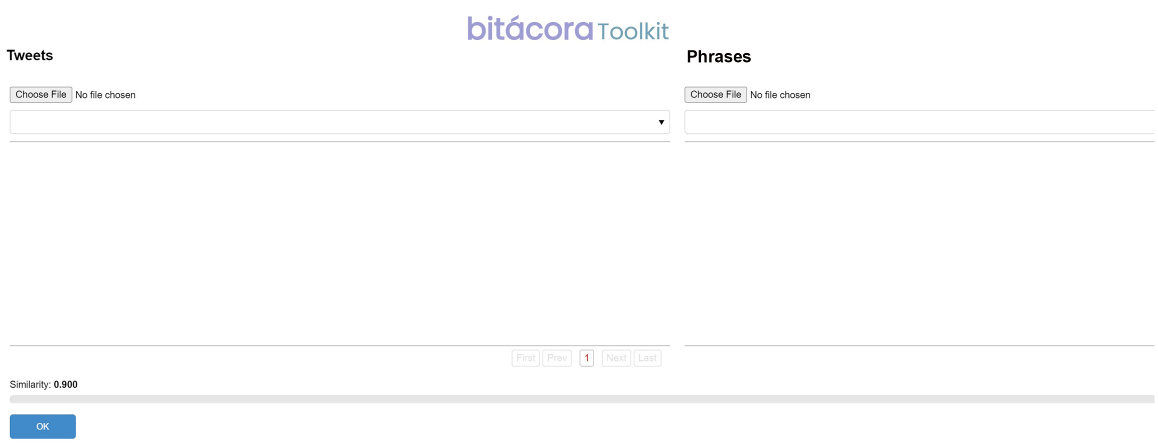


Figure 7.2: User Interface of Module 2

[2]Source code of Module 1 and Module 2 is available in the next GitHub repository: https://github.com/AdrianaAG

Measures to Prevent Misuses of the Toolkit

Because the toolkit enables accessing large amounts of data from Twitter using my research credentials of the Twitter API, we implemented a series of measures to avoid any potential misuses of the toolkit.

First, we implemented a system to control the access of users to the toolkit. For practitioners to access to Module 1 and module 2 they were required to access with a user name and password. I exclusively granted credentials to participants for the evaluation of the toolkit, and the credentials were restricted to the days for the days that I take the evaluation. In addition to the credentials, I restricted participants to the number of tweets they were able to retrieve per search.

A second precautionary measure we took was that the data downloaded by the tool only contained the Twitter handle and did not include the user name, this was done with the intention of avoiding downloading personal information.

A third measure was to warn participants during the introductory workshops of the evaluation of the toolkit about processing basic principles of personal data and I also emphasized that working with data from social media platforms implies interacting with communities, which require us to be respectful.

7.2.2 Overview of the Manual

The purpose of the manual is three-fold. First, it introduces the reader to understand the potential of social media data to inform the work of NGOs and civil organizations while cautioning the reader of the constraints and ethical implications of using such type of data. To this end, the manual includes case studies to illustrate how Twitter data has been used as a source of information to measure awareness, support decision-making on policies of various kinds, and encourage greater public participation. Additionally, the manual includes tools to guide the user on deciding whether Twitter data is the most appropriate source of information to inform their work.

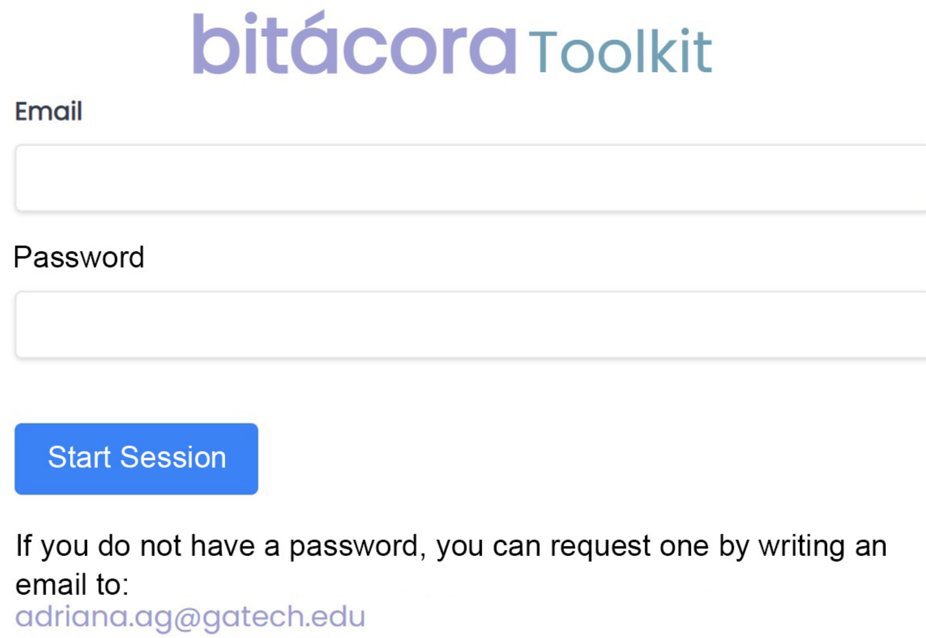


Figure 7.3: The interface of the login system

The second objective of the manual is to instruct the reader in the operation of the computational tools that make up the toolkit to search, collect, and analyze the tweets. Lastly, the manual guides the readers to interpret the data from Twitter with a critical and contextualized approach to integrating it as evidence in their work. To this end, the manual offers a series of lessons to accompany users on the process of defining a problem to explore on social media, evaluating whether social media is or not an appropriate source of data, and a guide to conduct two types of analysis, an exploratory and an in deep qualitative analysis.

The manual is divided into four sections:

1. **Introduction to the manual:** Provides a description of the toolkit bitacora, its main components, for whom its designed and for what contexts is it recommended to use.

2. **Introduction to social networks:** Includes case studies to inform about the strengths and limitations of data from social media.

3. **Analysis:** Section that informs about two types of analysis, the exploratory and in-deep analysis.

4. **Worksheets and qualitative tools:** Section with various worksheets to support and guide the user on framing their questions, scoping the problems to examine, and documenting their findings.

Next, I provide more detail of sections 2,3, and 4 of the manual, to illustrate how their design answer the last research questions of my dissertation:

1. **RQ 4:** What type of tools are needed to support practitioners in mobilizing social media data from online communities into their institutional context?

2. **RQ 4a:** How do we design tools that promote situated and critical perspectives when organizations practitioners integrate social media data into their work?

3. **RQ 4b:** What are the necessary guidelines to support organizations' practitioners in evaluating whether social media data is appropriate to inform their work?

4. **RQ 4c:** When mobilizing social media data to NGOs, how do we design supportive tools that document the context of the data's production?

7.2.3 Section 2: Introduction to Social Media Data

Before practitioners use the computational tools of the toolkit, it is essential for them to understand the characteristics, strengths, limitations, and potential risks that come with the integration of social media data into their work. The section two of the manual communicates the characteristics and strengths of social media data through four case studies that used content from Twitter and Facebook as primary source of data. These case studies illustrate how social media platforms have become new forms of collective organization, and they type of insights that is possible to draw from user-generated content. Furthermore, lessons of the case studies show that the content generated in social networks has the potential to: 1) identify community responses to local crises, 2) identify actors and actions that were not previously identified, 3) understand local narratives, capacities, and experience of citizens, and 4) differentiate how a phenomenon manifests in different communities.

In the manual, it is also described the limitations of social media data in terms of representation, explaining that data from social media are always incomplete datasets due to the fact that they only reflect the perspectives, opinions, and needs of a limited demographic sample, specifically those who use social networks and therefore are not fully representative. Additionally, in the manual, it is explained that due to the characteristics of data from Twitter, it is impossible to accurately determine the specific characteristics of the communities that generate data on the platform, such as profession, gender, or age group. The purpose of highlighting the restrictions of the Twitter data is to help practitioners set realistic expectations of what is possible to do with this type of data and encourage a more critical and cautious approach when integrating this content into their work. To simplify practitioners' decision process on whether or not to use social media data to inform their work, the manual includes a decision tree that makes it easy to rule out issues that cannot be examined with social media data. Lastly, the manual includes a page on the implications of collecting, storing, analyzing, and interpreting data from social media. While the ethical implications continue to be defined by the scientific community, I emphasize three principles that should be kept in mind by practitioners at any moment:

1. <u>Using social media data implies interacting with people and communities:</u> Behind each data set, there are communities that need to be respected. As researchers using data from social media platforms, it is necessary to recognize the existence of asymmetric dynamics between researchers and members of online communities. Since data from social media platforms are public and massively available, it is not strictly necessary to get informed consent, which implies that members of online communities are usually unaware of being monitored and lack the capability to decide which

of their data can be collected.

2. Social media data are only the starting point: It is imperative to be cautious of the type and reliability of inferences that is feasible to obtain from social media data. Depending on the tools and platforms used to collect data, some communities and perspectives become more visible than others. Regardless of the size of the data sets collected from social media platforms, they will always be biased by the demographic sample and geographic location, constraining whose experiences, needs, and interests are visible for analysis, affecting the inferences that we can draw from the data. For this reason, when integrating social media data into the context of NGOs, they should be used as starting point to be completed with data from other sources.

3. Follow the basic principles of personal data processing: When interacting with social media data, practitioners should consider following the basic principles of personal data processing to reduce harm and risk.

   (a) Data should be collected for specific and legitimate purposes. In this sense, the interface of module 1 allows users to search and explore Twitter data without necessarily downloading it. The intention is to motivate users to explore the data first instead of immediately downloading large datasets. Additionally, when downloading tweets with the toolkit the user names are not included.

   (b) After downloading the datasets with the toolkit, it is imperative to maintain the confidentiality of personal data at all times and do not distribute the toolkit datasets. It is also recommended not to keep the data for periods longer than necessary for the purposes for which they were collected. Lastly, erase the data after being analyzed.

   (c) To respect the rights of the participants, never use the exact content of the tweet and avoid referring to a specific account or user name. Instead, paraphrase the content of the tweet.

7.2.4 Section 3: Analysis

Section three of the manual describes the process for conducting two types of qualitative analysis —*exploratory and in-depth*— that allow practitioners to gauge expectations of questions that can be answered and issues that can be examined with Twitter data. This section of the manual is organized into two parts. The first section guides practitioners in an *exploratory analysis*. The objective of this initial analysis is 1) to define the problem in terms of the characteristics of the Twitter data, 2) to explore how the topic of interest is discussed on Twitter, 3) to determine if the use of Twitter data is relevant or not depending on the problem to be examined, and 4) assess whether the observed narrative is of interest to carry out a more in-depth study. The second section describes how to conduct an *in-depth qualitative analysis* by focusing on a small number of tweets corresponding to a specific period of time. As a result of this analysis, practitioners should be able to obtain evidence that enables more accurate inferences and informs potential next steps.

The purpose of providing two types of analyses in the manual is because practitioners first need to understand how a topic is being discussed on Twitter to tune the questions they seek to answer based on communities' interactions on Twitter. With this, my goal is to encourage practitioners to recognize the Twitter data limitations, challenging the notion that this data can address all types of questions.

7.2.5 Section 3: Exploratory Analysis

This stage consists of three steps: 1) definition of the problem, 2) initial exploration on Twitter, and 3) categorization of initial findings.

Step 1 - Definition of the Problem

The first step of the *exploratory analysis* is to define the problem to be explored in terms of the scope and focus of Twitter data analysis. The manual provides three qualitative tools to support practitioners in this process: 1) problem definition worksheet, 2) decision tree diagram, and 3) keyword documentation template.

It is recommended to complete first the *problem definition worksheet*, which allows practitioners to gather information about the context and communities involved in the problem they aim to examine, as well as document existing evidence and data that they might have. This worksheet intends to encourage practitioners to reflect on their goals and expectations for using Twitter data. Once practitioners have gathered more information about the problem in which they are interested in integrating Twitter data, it is encouraged to use the *decision tree* to evaluate whether or not Twitter is an appropriate source. If the outcome of the decision tree is positive, then practitioners should use the *keyword documentation template* to collect keywords that describe the characteristics of the problem and the communities and organizations involved.

Step 2 - Initial Exploration on Twitter

The second step of the *exploratory analysis* consists of doing an initial exploration of the problem on Twitter. Using the keywords, Twitter accounts and hashtags collected in the previous step, an initial search is carried out using the Module 1 with the aim of collecting evidence of what the problem being examined looks like on Twitter. The result of this exploration should be to determine whether or not the topic is discussed on Twitter, how it is discussed, and to identify the communities, organizations, or actors involved in the conversation. This step should be conducted iteratively, and with each iteration, practitioners ought to collect more keywords, Twitter accounts, and hashtags to inform, guide, and refine the following explorations.

In addition to Module 1, the toolkit includes the *exploratory search documentation* template to guide practitioners in the exploration process and help them to document the search results by considering the following aspects:

1. *Type of content:* The most frequently shared content on Twitter is a result of keyword search. For example, tendencies of what people tend to share the most such as news,

photos, videos, etc.

2. *Type of discourse:* Register observed tendencies in the discourse and discussions portrayed in the results by paying attention to whether the results returned tweets reflecting complaints, local actions, or any other example of collective organization.

3. *Actors involved:* Recording who the actors or communities discussing the issue are and the type of accounts that share relevant information on the subject examined. For example, pay attention to the accounts publishing or sharing information on the topic. Are these accounts from citizens, non-governmental organizations, the government, institutions, activists, journalists, etc.? In parallel, practitioners ought to reflect on the patterns observed. For example, actors that prevail in most of the results and those that are absent.

4. *Data availability:* Record the availability of the information found on the topic examined. The objective of the initial exploration is not to quantify the tweets on a specific topic but rather to make a brief inspection of the frequency of tweets that reflect the examined topic.

Additionally, the *exploratory search documentation* template enables practitioners to compare the preliminary findings obtained on Twitter with data from traditional sources. The goal is to encourage practitioners to reflect on whether there is any information available on social media platforms that is not reflected in typical data sources and try to identify the patterns and perspectives that prevail in the results and those absent.

Step 3 - Categorization of Initial Findings

The third step of the exploratory analysis consists of categorizing the initial findings. The objective is to associate the content of the tweets identified in the previous step with the definition of the problem being examined. To this end, practitioners need to review the findings of step two and define what type of tweets are emblematic or codify the dimensions of the investigated topic. As a result of this step, practitioners will be able to decide whether or not Twitter is an appropriate source of information for their purposes and if it is convenient to carry out a further exploration by analyzing the Twitter data in more detail.

The toolkit includes two work templates that guide practitioners through the process of categorizing initial findings —the Findings and Observations Documentation Template and the next Steps Documentation Template. The manual also explains that during the categorization, it is recommended to pay attention to the following aspects:

1. *Redundancy and information patterns:* In this context, redundancy should be understood as repeated information in the Twitter data. In some cases, observing redundancy in the data denotes information patterns, which can help confirm the actions of a specific group or provide new perspectives and insights on the topic being studied. Twitter data redundancy can take different forms, for example:

   - The type of organizations or communities that tweet the most about a specific topic.
   - The type of content that organizations share.
   - The population or community that tweets the most about a problem or situation, for example, activists, teenagers, government workers, etc.
   - The type of requests that a specific group demands on Twitter.

   Ultimately, the goal is to encourage practitioners to closely examine these patterns during their exploratory searches to define the topics and perspectives they want to delve deeper into and those they might want to avoid.

2. *Information not documented in traditional data sources:* As mentioned above, one of the strengths of Twitter data is that it reflects perspectives that are not always documented in traditional information sources. Therefore, it is recommended to evaluate

whether the data found on Twitter reveals actors, populations, or strategies of the affected communities that were not previously known or had not been considered.

7.2.6 Section 3: Qualitative Analysis

As mentioned before, the goal of conducting an in-depth qualitative analysis is to obtain evidence to be integrated into the NGOs' work. The steps of this type of analysis are: 1) data collection, 2) selection of data sets, 3) review and documentation of additional information, 4) categorization and 5) synthesis.

Data Collection

The first step is to collect and download the data using the module 1 tool, based on the findings obtained during the exploratory analysis. Like in the previous analysis, it is recommended to use the templates of *exploratory search documentation* and *findings and observations documentation*.

Selection of Data

The second step of the qualitative analysis consists of selecting a more specific Twitter data set to conduct an in-depth analysis. After collecting the tweets using the module 1 tool, practitioners ought to use the module 2 tool to identify tweets that contain content similar to the issue they are examining. It is possible that even after using the tool in module 2, the resulting number of tweets is still too high to do an in-depth qualitative analysis, for example, more than 1000 tweets. In that case, it is recommended to randomly select a reduced number of tweets for the in-depth analysis.

It is difficult to determine the exact number of tweets to be included in the qualitative analysis since it depends on the topic being examined and the time available for analysis. If practitioners have little time, it is suggested to select between 20 and 50 tweets per topic to analyze them in-depth. Alternatively, tweets can be filtered by selecting only those that include a word or term that refers to geographical areas, actors, keywords, or specific hashtags. This is an iterative step, and the objective is to reduce the number of tweets to a manageable number to perform an in-depth qualitative analysis.

Review and Documentation

The third step of the in-depth qualitative analysis consists of reading each of the tweets selected in the previous step and, in parallel, documenting additional information about the tweets to more accurately understand the context in which the content was generated. To this end, practitioners can use the *findings and observations documentation* template.

As previously mentioned, the objective of the templates is to record additional information about the tweets to guide practitioners in their analysis and interpretation. The additional information will depend on the examined problem. However, it is recommended to at least cover the following aspects:

1. *Actors involved:* Person or organization that publishes the tweet and the reactions it receives. When interpreting the content of the tweets is relevant to consider who publishes it (e.g., a citizen, an organization, a person from the government, etc.). Similarly, in the interpretation of the tweets, it is necessary to consider the reactions they have to assess their impact, for example, their number of likes or retweets.

2. *Documentation of patterns:* It is imperative to document the patterns observed, such as community problems, needs, responses, and collective initiatives reported on Twitter. Additionally, it is necessary to pay attention to the commonalities, distinctions, and relationships between actors surfaced by the analyzed content.

Categorization

The fourth step of the qualitative analysis consists of collecting evidence of how the examined problem manifests on Twitter. To this end, it is necessary to categorize the tweets previously defined and documented. The categorization of tweets should follow a classification defined by collaborators and counterparts involved in the project. The purpose is to interpret the data in a consensual way. After defining the categories, practitioners need to read and classify the tweets to identify patterns. The manual explains how to conduct this analysis using an affinity diagram, and like with any other qualitative data analysis, categorizing tweets is an iterative step and should be repeated until defining enough categories that satisfactorily describe the examined problem.

Synthesis

The fifth and final step of the in-depth qualitative analysis consists of synthesizing the data previously collected and categorized. The goal is to make sense of the collected tweets by looking for patterns and correlations in the data to answer the initial questions that prompted using Twitter data in the first place. The output of this step should be a characterization of the examined problem. As mentioned before, data from social media platforms is always incomplete due to its limitations in terms of representation. Therefore, the observations collected through Twitter will never be exhaustive on the examined social issues. However, data from Twitter can offer evidence of collective reactions, responses, and perspectives on social issues that are otherwise very difficult to identify. Therefore, due to the limitations and restrictions of the data, the manual suggests that the most appropriate way to make sense of the data collected on Twitter is to follow a holistic approach to interpreting the data and understanding the examined problem. This holistic approach can take various forms, but the main idea is to characterize the different components such as actors, geographies, perspectives, and themes that outline the examined problem. Thus, the characterization is linked to the initial definition of the problem. Additionally, the manual provides the following suggestions to organize the data:

1. *Geographic areas:* organize the findings considering the location of the actors involved. If the data allows it, focus the analysis on the particularities, differences, ca-

pacities, and needs of communities, depending on their geographical location, which can be by state or city.

2. *Type of solution:* Solutions that respond to local needs while surfacing the communities assets.

3. *Action - situation:* Observations or findings exemplifying a collective action or response to a particular situation.

4. *Local problems:* Findings exemplifying problems rooted in the communities.

In conclusion, regardless of the organization of the analysis, it is recommended to follow an approach focused on identifying perspectives, capacities, and knowledge within the community that was not considered before while answering the initial questions that motivated the study.

7.2.7 Section 4: Description of the Worksheets

The purpose of designing this set of worksheets is to provide participants with tools that allow them to document their decisions in exploring and integrating social media data into their work. I designed five worksheets to be used along with the platform and any of the two proposed types of analysis.

Worksheet 1 - Problem Definition Template

This worksheet guides practitioners in defining the problem to be examined using data from Twitter, and it consists of five sections (Figure 7.4).

1. **Description and expectations:** This section is at the center of the template, and it asks for a short description of the problem and users' expectations using data from Twitter.

2. **Context (place and time):** This section asks three questions: 1) where is this problem happening?, 2) is it a widespread and generalized problem or a problem of a specific area?, and 3) what is the timeframe of the problem? The intention of including these questions is to encourage the user to reflect on what information they have about the problem to be examined and the problem characteristics in terms of place and time.

3. **Communities:** This section asks for the communities involved in the problem to be examined.

4. **Evidence and data:** This section asks the following questions: 1) what evidence is currently available on the problem? 2) what kind of evidence do you expect to find? 3) what type of data is currently available?

5. **Objectives:** What is the motivation for using Twitter data? List the questions that you seek to answer with the exploration of Twitter

Worksheet 2 - Keyword Documentation Template

The objective of this template is to document the keywords that characterize the problem that practitioners want to explore on Twitter. The template has five columns of possible keyword categories that users might decide to register. The categories are:

1. Description of the problem

2. Names of organizations involved

3. Affected communities

4. Twitter accounts of actors involved

5. Hashtags

Worksheet 3 - Exploratory Search Documentation Template

This template is to help users document their observations from exploratory searches. In the manual, I suggested practitioners use the worksheet to keep track of which combinations of words yield the most unusual or appropriate content that contributes to the understanding of the examined problem. The template is divided into seven columns with the categories of aspects that I recommend to document during the first observations.

These categories are: 1) place, 2) time, 3) the combination of keywords, 4) number of results, 5) observations, 6) actors/Twitter accounts, and 7) examples of tweets.

Worksheet 4 - Findings and Observations Documentation Template

The purpose of this template is to document and synthesize the most relevant conclusions of the initial examination. Ideally, this template needs to be completed after analyzing the findings of the initial explorations on Twitter in the worksheet 3. On the left side of the template, there are three small boxes where practitioners can enter the project's name, a brief description, and a summary of the focus of the exploration, for example, listing some of the search words. The purpose of these three boxes is to summarize the project. In the template, there is also a text box divided into three sections:

1. **List relevant actors:** This section asks to list any new actors or communities involved in the problem examined that might have been found on the exploration of Twitter. The form asks 1) who are they? 2) what is interesting about them? 3) what is their role?

2. **List initial findings:** This field asks to describe how the topic of interest is discussed on Twitter and recommends recording examples of tweets that are illustrative of the findings, as this facilitates the communication of results.

3. **List relevant actions or practices:** This section asks practitioners to document any

new identified perspectives or interesting initiatives that contribute to the problem examined and to consider answering the following questions:

(a) What is the purpose of these initiatives?

(b) What distinguishes them from existing initiatives?

Worksheet 5 - Next Steps Documentation Template

The purpose of this template is to encourage practitioners to reflect on the observations collected in the exploratory searches, assess whether the narrative observed on Twitter is of interest for further analysis, and determine the next steps. After participants complete the template, they would need to decide whether or not it is advisable to continue using data from Twitter to report the problem examined. Similar to worksheet 4 on the left side, there are three small text boxes where participants can synthesize the project focus and the keywords they used during the exploration. The template consists of the next three sections:

1. **Similarity between Twitter findings and prior knowledge:** In this section, the template asks practitioners to reflect about 1) how are the observations found on Twitter related or different from what you knew about the problem before?, 2) what's new?, and 3) what information or patterns are repeated?

2. **Potential limitations for the next steps:** This field asks practitioners to consider the constraints of integrating data from Twitter into their work. The template asks:

   (a) What limitations do you anticipate? For example, not finding information about a specific place or population.

   (b) What are some questions or concerns?

3. **Evaluate whether an additional search is or not necessary:** This section asks practitioners if they consider it necessary to carry out a more in-depth search and analysis.

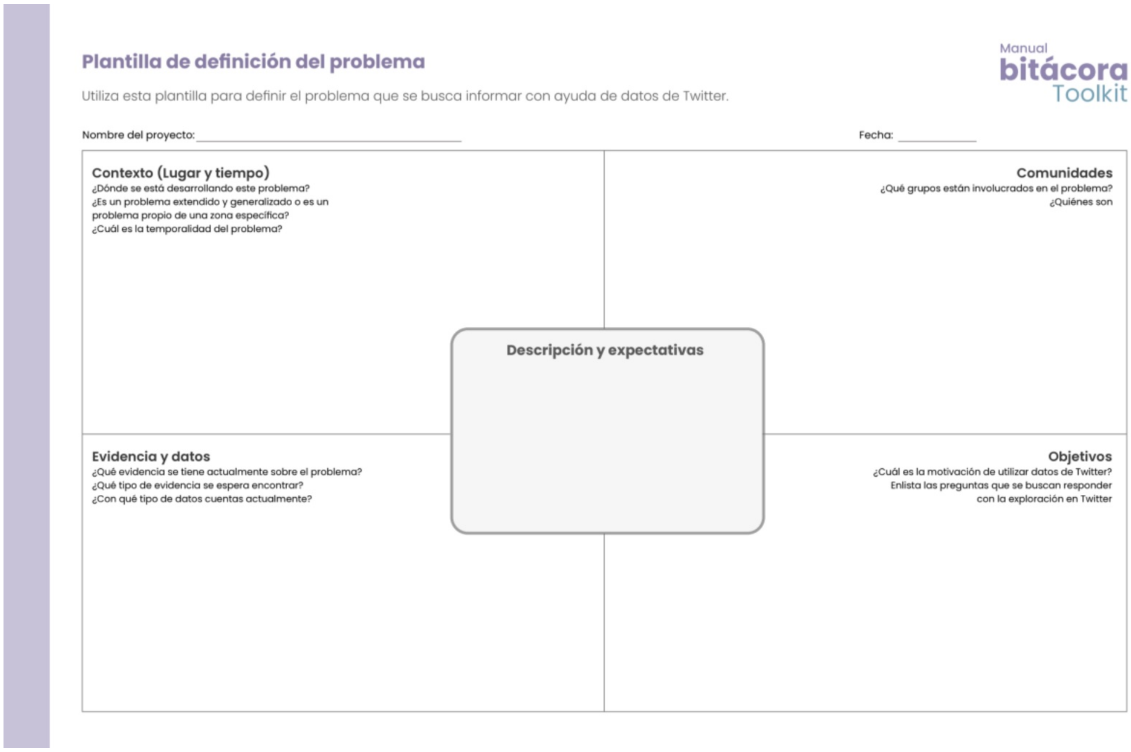


Figure 7.4: Worksheet 1 - Problem Definition Template

(a) If the answer is yes: How do you think Twitter data contribute to understanding the research problem you are examining?

(b) If the answer is no: What would be the reason?

Plantilla para documentación de palabras clave

Manual bitácora Toolkit

Utiliza esta plantilla para registrar las palabras clave que caracterizan el problema que se desea explorar en Twitter.

Nombre del proyecto: ____________ Fecha: ________

Descripción del problema	Organizaciones involucradas	Comunidades afectadas	Cuenta de Twitter de actores involucrados	Hashtags

Descripción:

Figure 7.5: Worksheet 2 - Keyword Documentation Template

Plantilla para documentación de búsquedas exploratorias

Manual bitácora Toolkit

Utiliza esta plantilla para documentar tus observaciones de las búsquedas exploratorias

Nombre del proyecto: ____________ Fecha: ________

Documentación de búsquedas exploratorias

ID	Lugar	Temporalidad	Combinación de palabras clave	Número de resultados	Observaciones	Actores / Cuentas de Twitter	Ejemplos de tuits
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

Descripción:

Figure 7.6: Worksheet 3 - Exploratory Search Documentation Template

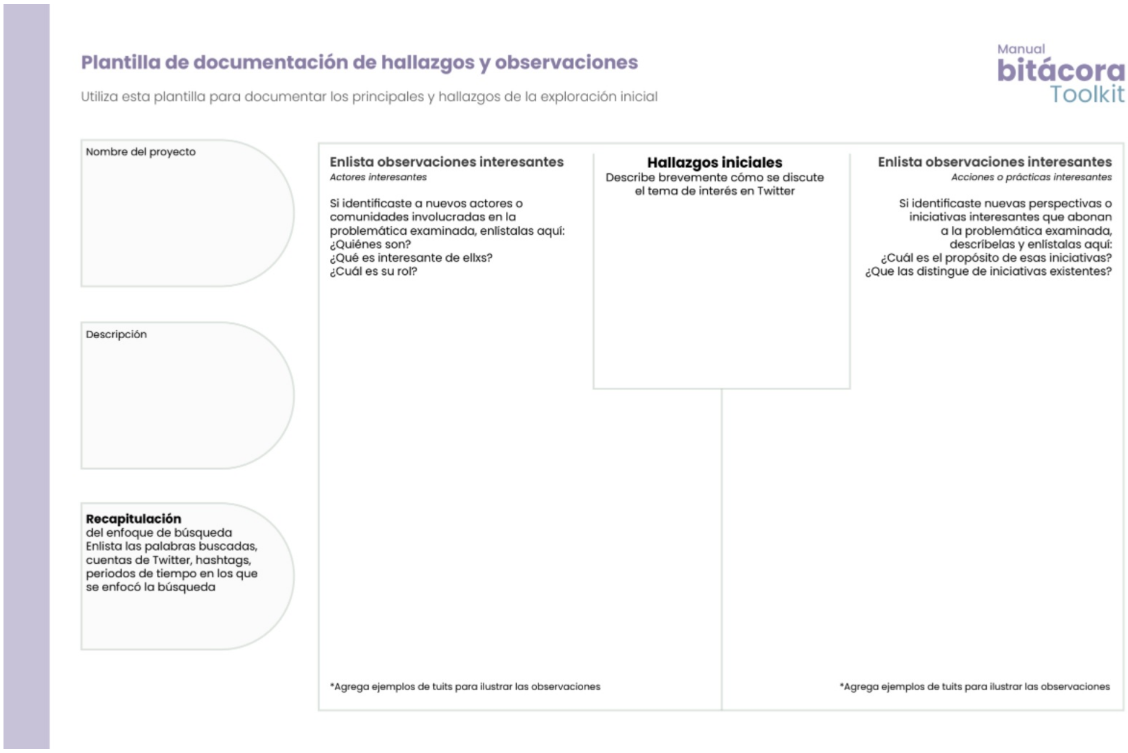


Figure 7.7: Worksheet 4 - Findings and Observations Documentation Template

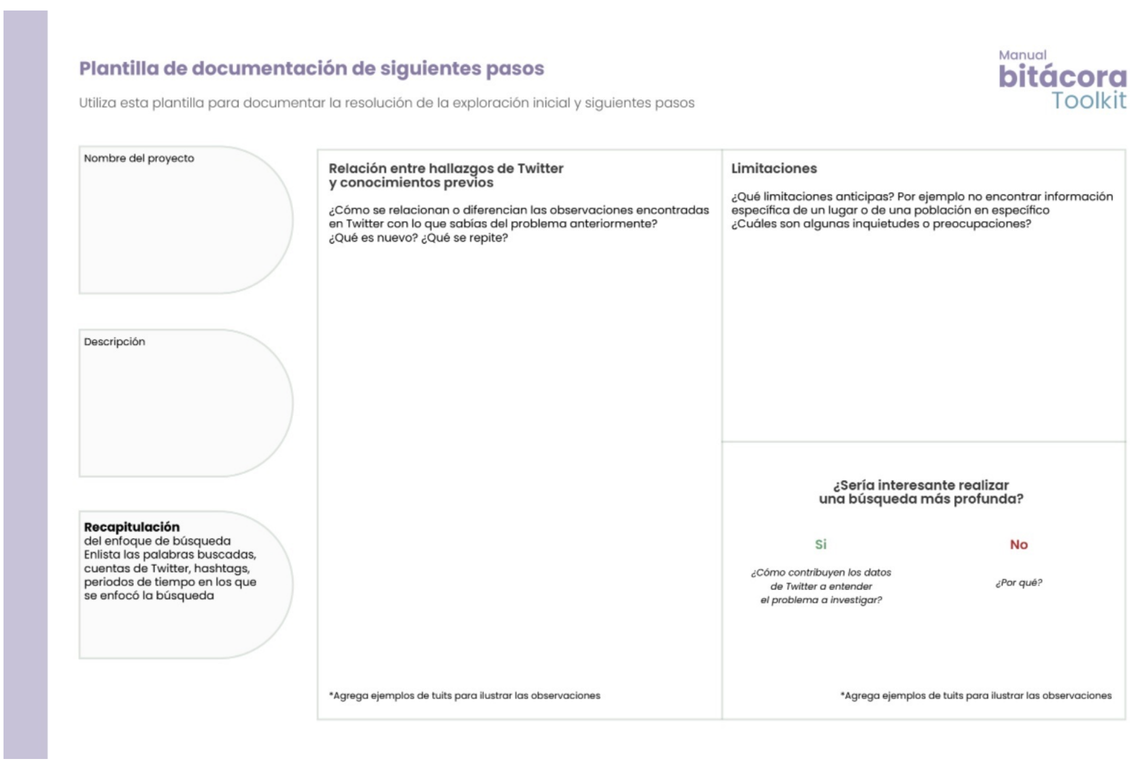


Figure 7.8: Worksheet 5 - Next Steps Documentation Template

CHAPTER 8

EVALUATION OF THE TOOLKIT

8.1 Overview of the Evaluations

Both the epistemologies of objectivity of those doing the work of data to determine what counts as evidence and the tools used to analyze the data shape the process of mobilizing digital traces from Twitter or any other social media platform into the institutional context of organizations that might benefit from the data. In contrast, existing scholarship on Social Computing has focused mostly on offering technical solutions to leverage social media data with an abstract approach without considering the social context of the communities that aim to inform. The account that follows in this section is a situated evaluation of the toolkit. I appeal to the notion of *situated* from Haraway, who advocates for surfacing the multiplicity of local knowledges and approaches that motivate "translating knowledges among very different –and powered differentiated– communities" [212]. Conducting a *situated* evaluation of the toolkit offers an opportunity to expose the features and limitations of the manual and the platform, along with the different ways the toolkit might affect participants' perception of the capabilities of data from Twitter.

The assessment of the toolkit consisted of two evaluations studies. The first study was conducted with members of the Accelerator Lab and took place in the last week of February and the first week of March. The second evaluation was conducted with two non-profit organizations in Mexico, SocialTIC and CIEP, through an ongoing collaboration of both organizations with the Accelerator Lab. The sessions of this evaluation took place during March. Both evaluations aimed to assess the toolkit in guiding practitioners to scope the problem in terms of the data available from Twitter, to design the search of tweets, and to evaluate whether or not Twitter was an appropriate source of information to complement their work.

8.1.1 Evaluation with the Accelerator Lab

The assessment of the toolkit with the Accelerator Lab consisted on conducting a one-week evaluation comprising two group workshop sessions with the Accelerator Lab Team, and one individual session, in which participants interacted with the toolkit by themselves. Each of the group session lasted between an hour and an hour and a half. All sessions took place online via Zoom and the activities were conducted in parallel on the visual tool MURAL. Six participants attended to the first session —four members of the Accelerator Lab team and two interns— and five participants assisted to the second session. After the one-week evaluation, I conducted an interview with each participant. The group session allowed participants to brainstorm ideas, exchange strategies, and listen to each other's questions. Then, the individual sessions allow participants to spend time with the toolkit without additional guidance other than the manual and the lessons from the introductory workshop. Lastly, the interviews gave me a chance to gain a more detailed understanding of participants' experiences and expectations on the toolkit and the data from Twitter.

To document the activities, we used MURAL, a collaborative workspace that provides digital boards. In MURAL, I created an individual workspace for each participant with the five worksheets of the manual. Each participant received credentials to use the platform, and they had access only during the week of evaluation of the toolkit. Additionally, the number of tweets they were able to download was restricted. Participants were also continuously reminded not to share the credentials with anyone else. Before the first workshop session, participants received an email with the digital version of the manual, the link to MURAL, and their credentials to access the platform.

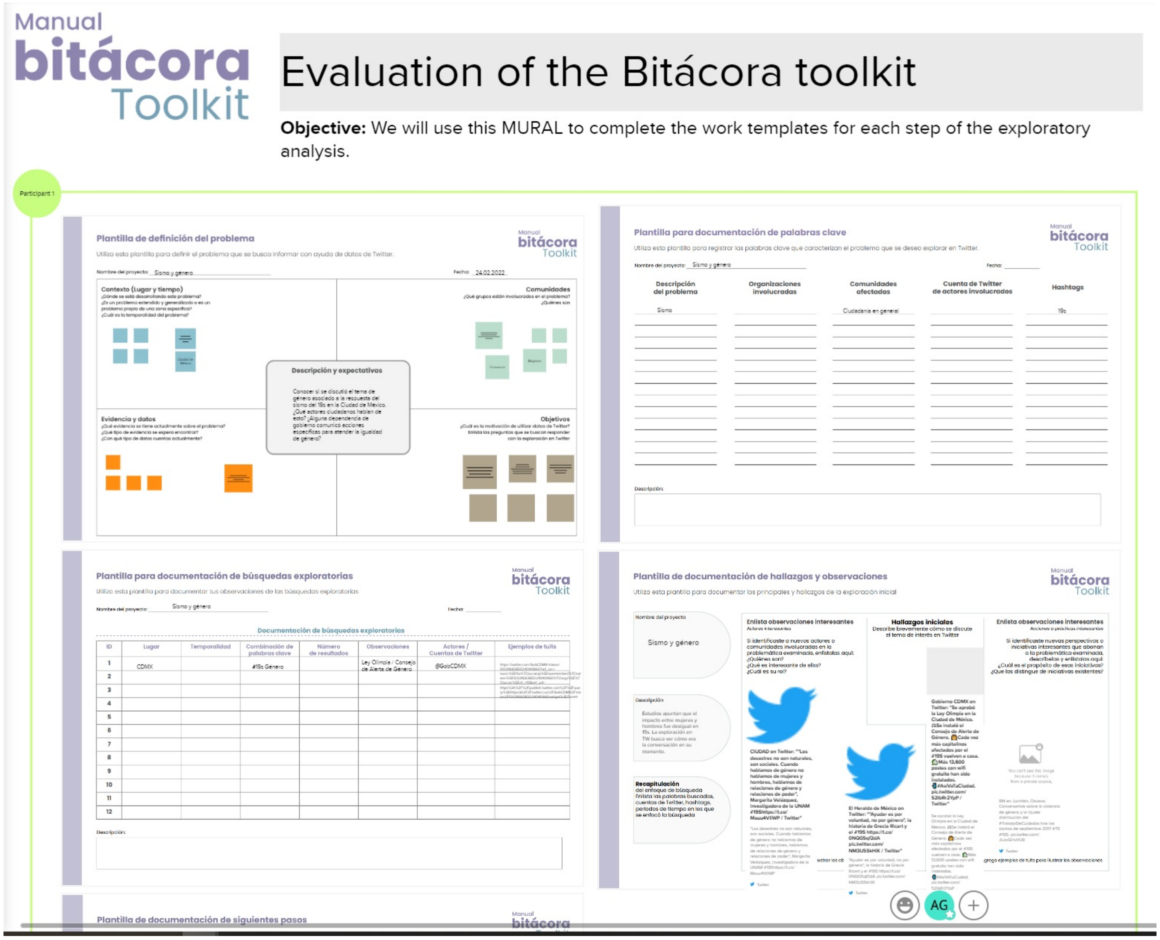


Figure 8.1: An example of the workspaces created in MURAL for the first workshop

Table 8.1: Breakdown of activities during the first session of evaluation

Time	Activity
15 minutes	Introduction to the toolkit
10 minutes	Overview of the evaluation of the toolkit
10 minutes	Implications of using social media data
15 minutes	Each participant completed the *problem definition worksheet*
10 minutes	Participants shared their topics to examine using Twitter
5 minutes	Conclusion and final remarks

First Group Session

The objective of the first session was to introduce participants to the components of the toolkit, the timeline of the evaluation, the implications of using social media data, and guide them in defining a problem to examine social media data using the toolkit. Table 8.1 provides a breakdown of the activities.

At the beginning of the session, I described the motivation and purpose of the toolkit, the content of the manual, and showcased the platform. During the showcase, I highlighted the main functionalities of module 1 and module 2, the elements of the interface, and some general considerations when using the toolkit.

After introducing the toolkit, I described the activities of the evaluation that they would need to complete individually and in the group. Specifically, I explained the role of the worksheets in conducting the exploratory analysis of the data from Twitter. At this point, I also highlighted that the purpose of the toolkit is to facilitate and guide users in conducting a qualitative examination of the data.

To make participants aware of the implications of using social media data, I briefly described the limitations of social media data in terms of representation and the kinds of problems that can and cannot be examined using social network data. Also, I provided participants with examples of situations in which data from Twitter has provided insights into how communities deal with local crises. I concluded this portion of the session by highlighting three basic principles of personal data processing for them to consider while interacting with data from social media.

Once we concluded with the introduction to the toolkit and discussed the capabilities of social media data, we moved to the MURAL, where every participant had their board with the five worksheets included in the manual. The purpose of having the worksheets in MURAL was to provide participants with a workspace to document their experience using the toolkit. Participants spent ten minutes completing the first worksheet – *problem definition template*– in MURAL to define the problem they wanted to examine with the toolkit. To conclude the session, each participant briefly described their ideas. However, due to time constraints, we could not discuss in detail each of the problems, so I sent each participant an email with suggestions and considerations for them to define the problem.

Individual session

The purpose of the individual session was to give participants the opportunity to interact with the toolkit only guided by the manual and the information provided to them during the first session. The manual provides two types of analysis of data from Twitter. First, an *exploratory analysis*, which, as I explained in the previous chapter, aims to provide readers with an initial understanding of the type of problems that are discussed in Twitter. The main tasks of this exploratory analysis are:

1. Defining the problem in terms of the characteristics of the Twitter data
2. Exploring how the topic of interest is discussed on social networks
3. Determining whether the use of Twitter data is pertinent depending on the problem to be examined.

Therefore, for the individual session, participants were asked to carry out an exploratory analysis completing the three steps included in that analysis, and at least complete the last worksheets of the manual, which is the *Next Steps Documentation Template.*

Second Group Session

Out of the six participants from the first workshop, only five participants attended the second group session, given that one member of the Accelerator Lab had to be absent due to a personal situation. The objective of the second group session was to let each participant report on their experience using the toolkit and present the two final worksheets of the exploratory analysis, which are the *findings and observations documentation template* and the *next steps documentation template*. The guiding questions for the discussion were:

- What strategy did you follow to search for the tweets?
- What were your conclusions and findings after conducting the exploratory analysis?
- How did your understanding of the problem change after analyzing the tweets?
- How would you describe your current expectations from the information on Twitter?
- Would you conduct an additional search?

Due to time constraints, we could only spend between ten and fifteen minutes discussing the findings and experience with the toolkit of each participant. Therefore, I conducted follow-up interviews with each participant to learn more about their interaction with the toolkit.

Interviews

I interviewed the participants who participated in the second session, and each interview lasted between thirty minutes and one hour. All the interviews were conducted in Spanish, audio-recorded, and then transcribed and analyzed. The objective of the interviews was three-fold. First, gather detailed feedback on participants' experience using the manual and the platform. Second, gain a deeper understanding of participants' initial expectations using Twitter and if they changed while using the toolkit and how. Third, to better understand how participants would imagine integrating the toolkit into their everyday work and the potential limitations and constraints on doing that, I asked participants about their tasks in the Accelerator Lab and their previous experience conducting qualitative research.

8.1.2 Evaluation with NGOs

The second evaluation was conducted with two non-profit organizations, SocialTIC and CIEP. The Accelerator Lab is currently working with both organizations on the challenge of improving public consultations. For this reason, one member of the Accelerator Lab accompanied me in each of the sessions, helping with moderating the discussions and providing feedback on the design of the activities. Our motivation for conducting a second evaluation with two civil organizations was twofold. First, examine what could be the potential value of data from Twitter to organizations with such a different approach than the Accelerator Lab –which focuses on sustainable development. Second, gain an initial understanding of what would entail for the Accelerator Lab to adopt and socialize the toolkit with civil organizations. Due to time constraints, it was not possible to conduct interviews with participants of the second evaluation session. However, the evaluation continues and will be completed in the following months.

Overview of the Organizations

SocialTIC is a civil society non-profit organization that aims to empower agents of change in Latin America by reinforcing their analysis, social communication, and advocacy actions through the strategic use of digital technologies and data. Their focus is to train and accompany groups and individuals in info activism, the use and opening of data, and digital security. Members of the organization promote multidisciplinary learning and collaboration spaces, research and develop new tactics and tools, and encourage other Latin American groups to generate new projects and replicate training actions locally. On the other hand, CIEP (by its acronym in Spanish Center for Economic and Budgetary Research) [1] is a non-profit civil society research center that provides information and analysis to influence, improve, and democratize discussions and decision-making in the economy and public finances sector. The organization focuses on diverse areas of public finance, such as public debt, public spending, and income and taxes, among others.

[1]In Spanish: Centro de Investigación Económica y Presupuestaria.

Logistics

The evaluation consisted of a similar plan to the previous one with one additional group and individual sessions. Therefore, the assessment consisted of a two-week evaluation involving three workshop sessions and two individual sessions. The activities of the first workshop session were the same as in the evaluation with the Accelerator Lab. The first half of the second session consisted of sharing findings of the initial exploration, and in the second half, I introduced module 2. The third group session consisted of debriefing their experience using the toolkit and defining what could be the potential steps for them to capitalize on the findings found on Twitter.

Each group session lasted an hour. All sessions took place online via Zoom and the activities were conducted in parallel on the visual tool MURAL. Two members of SocialTIC and three members of CIEP organization attended the group sessions.

Similar to the previous evaluation, the group session allowed participants to brainstorm ideas and listen to each other's questions regarding using the toolkit. During the individual sessions, participants of the SocialTIC organization worked together, while participants of the CIEP organization worked individually.

Like the first evaluation session to document the activities, I created an individual workspace in MURAL for each participant. The MURAL workspace used for the first session was the same we used in the evaluation with the Accelerator Lab. However, for the second and third sessions, we used different MURAL workspaces to support the activities of those sessions.

Participants received platform credentials that were active during the evaluation, and the number of tweets they were allowed to download was restricted too. Participants received the same recommendations I made during the first evaluation regarding not sharing their credentials with anyone else. Before the first workshop session, participants received an email with the digital version of the manual, the link to MURAL, and their credentials to access the platform.

Third Session

The first two group sessions were barely enough to help participants get familiar with the toolkit and gather initial findings using the Twitter data. As the evaluation progressed, it became evident that a longer engagement was needed to integrate participants' findings into their organization's work. Therefore, the third session consisted of defining with participants the following steps to capitalize on the data from Twitter.

I started this session by gathering feedback on participants' overall experience after interacting with the toolkit. In the session, we had three activities where we first went over the initial worksheet where participants recorded the definition of the problem they wanted to examine with the toolkit. Then, in the second activity, we brainstormed together on how to turn the observations collected on Twitter into actionable information by reflecting on what we learned, what was missing, and how we might go about it. Lastly, in the third session, we set goals and began outlining specific mechanisms to move forward.

8.1.3 Data Analysis

All the workshop group sessions and interviews were conducted in Spanish, video recorded, and transcribed. I gathered six and a half hours of data from the five workshop sessions and four hours of data from the interviews. Additionally, I analyzed the MURAL materials, such as the worksheets completed by participants. In reporting the findings, I anonymized the names and genders of participants to protect their identities and translated to English the quotes included here. To analyze the data, I inductively and iteratively coded the transcripts using memoing and coding.

8.2 Findings

8.2.1 Lessons on the Process of Seeking Valuable Information on Twitter

All the participants decided to examine topics related to ongoing projects or inform future projects developed in the lab. The objective of participants' examinations fell into any of these three categories:

1. Investigate if a problem was present in the public sphere of Twitter, and examine the characteristics of the discussion

2. Identify which actors were involved in discussing the topic of interest (e.g., government, citizens, civil organizations, and the public sector) to understand their discourse and position

3. Gather evidence to decide on potential places of fieldwork

In the initial brainstorming, participants framed the problems to examine on Twitter very broadly and covered various topics such as resilience and gender, community response to natural disasters, mining companies in Mexico, gentrification, consultation with indigenous communities, and access to water and droughts. Due to time constraints in the first session we could not deepen the discussion so after the initial session, I sent them individual feedback and recommendations on how to scope their exploratory analysis, potential limitations, and suggestions on how to collect initial keywords. Table 8.2 provides a synthesized description of participants' topics, initial framing of the problem, and the framing that participants ultimately followed.

Table 8.2: Breakdown of the topics participants of the first session explored

ID	Topics	Initial Framing of the Problem	Final Framing of the Problem
1	Resilience, risks and gender	How do the government and civil organizations discuss these topics?	Examine if there were discussions around gender disparities in response to the earthquake in Mexico in 2017. Who talks about this? Did any government agency develop target actions to address gender inequality?
2	Community response to natural disasters	Catalog of practices and responses of organizations and government for attention or recovery to communities after a natural disaster. Choose five iconic cases where there has been a catastrophe in the country.	Examined and compared the citizen organization and communication in two crises due to natural disasters. The participant examined the response to the earthquake in Mexico City in 2017 and the landslide in 2021 in the Cerro del Chiquihuite located north of Mexico City.
3	Mining companies in Mexico	Recently, the Supreme Court of Mexico revoked concessions to Canadian mining companies. What is being said about it?	Narrowed the search following the feedback received and focused on searching information about a community in the state of Puebla where the mining concession of a Canadian company was revoked.
4	Gentrification, increased rents in Mexico, and displacement of local people by foreigners	Examine if the increase in the cost of rent is happening systemically and if it has the potential of becoming a social problem.	Narrowed the scope of the search to examine the trends of rental prices in Mexico City.
5	Consultation with indigenous communities	Examine if actors from the private sector or government had discussed this topic and their respective position. A potential limitation is that indigenous communities living in rural areas might have restricted access to the internet.	Reduce the scope of the search to only examine the opinion of private sector actors regarding indigenous communities.
6	Access to water and droughts	Lack of water in Tepoztlán, Morelos. What is being said on Twitter about the lack of water in Tepoztlán? Evaluate if it could be a location for a future project. A potential limitation is that people living in Tepoztlán might not necessarily use Twitter.	N/A

The feedback I gave to the participants aligned with the guidelines I proposed in the manual. My recommendations focused on narrowing the topic, the timeframe of the search, and identifying keywords through other sources such as newspaper articles or by searching for civil organizations focused on the topic they were examining. For example, for the participant that analyzed the discussion on Twitter about the Canadian mining companies in Mexico whose permits were revoked, my suggestion was to search for news about the decision of the supreme court to identify 1) when the decision was made public, 2) the name of the Canadian companies, 3) the location of the companies and 4) communities impacted by those companies. While my recommendations aimed to reduce the search scope, according to the experience of some participants to find interesting insights on Twitter, what was needed was to be more flexible in structuring the search and be willing to explore.

> *One of the things that I noticed with the rest of the participants is that they did not have many results because their problem was already very defined [...] and precisely what helped me was to search for data outside of the timeframes of the main event I was examining, and I found something relevant related to the earthquake and gender that occurred two years after the earthquake [...] It is not that you have to define too many variables from the beginning. It is more like a problem is pointing in a direction, so I am going to try all these combinations.*- Avellaneda, Accelerator Lab member

The recommendations in the manual guided practitioners to define keywords, timeframes, and location on the presumption that after the event of the study has happened, the discussions on Twitter are immediately triggered, and after a couple of months, the conversation fades out. However, some of the most relevant findings that participants reported during the second session came from discussions that happened in unexpected time frames, as described by Avellaneda.

> *Something I didn't expect was that a critical moment in the gender and re-*

silience conversation happened two years later [after the earthquake], in 2019, when the university started publishing a series of research studies that provided more evidence of the unequal impact of the earthquake between women and men. In parallel to the publication of the studies, there was a wave of tweets and newspapers promoting this research and discussing gender inequalities. So, the fact that this discussion happened two years after the earthquake was very interesting to me.- Avellaneda

Here Avellaneda is reporting that they found relevant information to what they were searching for in tweets and conversations that happened two years after the timeframe they were focused on, which was September 2017, when the earthquake happened in Mexico City. Avellaneda's comment surfaces the assumptions contained in the manual, specifically that the discussion of topics happens linearly. Instead of defining a specific timeframe for searching tweets, Avellaneda decided to search with a wide timeframe and explore the results searching for patterns and insights.

When searching data on Twitter on social problems that do not fall into the definition of crisis (e.g., natural disasters), it is more difficult to determine a precise time frame. However, crises can trigger discussions and reveal the seriousness of social problems neglected for a long time. For example, Lucas was interested in understanding gentrification and the seriousness of the situation in Mexico, but due to the nature of the problem, it was difficult for them to define a precise time frame to collect data. To address this, Lucas searched for tweets using a wide time frame and found that the conversation spiked during periods of crisis.

I found tweets from people who complained years ago and others who complained last year and noticed moments in time when the topic became a trendy conversation on Twitter. For example, when the earthquake happened, it was a moment when rents went up because there were a lot of people displaced that were left homeless and had to find another place to live. During that *time, there was a lot of demand for housing, and landlords started raising the prices. I noticed the same trends during the pandemic.*-Lucas, Accelerator Lab member

Another recommendation in the manual refers to use as a starting point organizations and actors' names that are recognized as agents involved in the subject to be examined. This recommendation assumes that the topics explored on Twitter by practitioners will be well-defined and with clear actors. However, when participants first brainstormed about the topics they would like to examine on Twitter, they turned out to be too abstract or difficult to describe just using some keywords. Participants were interested in examining development problems, which despite always being present in our society, are not necessarily topics that people discuss on Twitter using the same terms they use. As Lucas's quote shows, the members of the lab were conscious about the lack of discussion of these topics in Twitter.

I don't know if there's like a conversation about this, because honestly what we do in the lab, I mean these are not topics of conversation on social media [. . .] so, sometimes it is like saying well, we know that the conversation always goes in another direction that has more to do with what all the organizations are doing, so let's look for what the conversation that already exists to also know how an issue can be addressed.-Lucas

Similarly, Avellaneda mentioned changing their search strategy because their topic of interest was not necessarily something people would be discussing at the time the search was conducted because there was not an ongoing crisis in Mexico City.

I changed well I readjusted the topic that I initially proposed. I was previously interested in looking at issues of resilience and gender. But we are not in a moment of crisis, and on Twitter, people tend to post what they have at the top of their minds, and I thought it would be difficult to find something interesting.

So, I decided to focus the search on the earthquake of September 19, 2017, and investigate the relationship between the earthquake and the unequal impact between women and men.- Avellaneda

Avellaneda's strategy to examine their topic of interest while considering how conversation trends change on Twitter during times of crisis, lead them to develop an analysis strategy that consisted of searching for *analogous cases*.

For example, if a flood happens and I am interested in examining how citizens responded to it, maybe it would be helpful to find out how citizens responded to an earthquake that occurred in the past. And this would let me know what I could expect to see or what I should look for while looking into the flood. Because if I search about the flood using the key words first aid or using the dates of the flood, I might not find anything. But if I then search for an earthquake and find out that when people talked about the earthquake, they didn't talk about first aid but instead talked about human rights. Maybe I can go back to the flood and think about human rights. I think using analogies to examine a problem is interesting because it may not offer us the result we imagine but brings us closer to understanding the problem in another way.

The fact is that sometimes we are not going to find those exact words that can bring us precisely closer to our search because we do not know what is going to bring us closer to the answers because we do not know the shape of the answers. So maybe examining how problems evolve in other contexts can give us some insight.- Avellaneda

8.2.2 Social Media Data as a Staring Point

The toolkit was useful in helping participants find information on Twitter that contributed to their understanding of the problem, either by identifying other actors or new evidence that informed the dimensions of the problem studied.

Findings from the initial exploration allowed Rolando to noticed the differences between the crises he was examining – the earthquake and the landslide– in terms of citizen organization, collective discourse, amount of tweets and actors involved. Regarding community response, Rolando noticed that people were using the hashtag #Verificado19S on Twitter as a way to cluster people's needs and to create an inventory of what it was being needed and where, in real time. Although Rolando's analysis focused on the first week after the earthquake, which could be considered a short time, he was able to identify changes on citizens' communication practices and needs. He observed that the majority of tweets that were posted right after the earthquake showed that people were requesting for phone access, and then the needs of the people were rapidly changing.

> *It was interesting to see that the hashtag #Verificado19S was a way to communicate what people needed across the city. The hashtag helped people coordinate picking up and delivering supplies and updating the inventory in real-time and organically. With Twitter, this supply chain became more efficient. As time went by, people stopped using the hashtag #19S and began making other types of demands, especially petitions to the secretary of civil protection.*- Rolando, Accelerator Lab intern

Ronaldo also noticed a difference between both crises in terms of the numbers of tweets found with the toolkit. While several tweets discussed the earthquake, few reported on the landslide, and most of them were either what Rolando called *informative tweets*, meaning tweets with links to news, or tweets with a political tone e.g. criticizing the government. According to Rolando, the tweets about the landslide had a political tone because three months before the landslide, the president closedown the Natural Disaster Fund (FONDEN), a program created in 1996 to help people affected by natural disasters; this made the discussion focus on political issues and not on the crisis itself. Therefore, due to the small number of tweets and their political tone, it was difficult to get any insight into citizen and government responses after the landslide.

> *After collecting the tweets, I skimmed them and found the word FONDEN, the Fund for Natural Disasters, and the president shut it down with a decree. The tweets about the Chiquihuite (landslide) were primary informative tweets, for example, reporting how many people had died and how many people were affected, that kind of tone. I think the tone of the tweets was so political because people considered that the president was in part responsible for the impact of the landslide. He shut it down, and precisely in these crises, it became more evident that those funds were necessary.*- Rolando

The type of information available on Twitter will always depend on the context of the crisis. While the tweets about the earthquake revealed the collective efforts of citizens to aid others in need, the tweets about the landslide did not offer any evidence of government or citizen organization. Instead, they revealed the dissatisfaction of the citizenry regarding the shutdown of the FONDEM. Moreover, the relevance of the findings from Twitter depends on the NGO's objectives when conducting the analysis. In this case, the results obtained from the landslide were not useful to the participant because his goal was to collect the communities' practices and responses to recover after a natural disaster. However, this confirms the need to conduct always an initial exploration of Twitter to evaluate whether it is not a valuable source of information

Participants also reported that their examination of the data from Twitter allowed them to reach conclusions that extend, and in some cases changed, their initial understanding of the problem examined. For example, Lucas, who researched the topic of gentrification in Mexico City, began their search with the assumption that foreign people coming to Mexico City were causing a rise in the rent prices. However, based on some findings from Twitter indicated that real estate was partially responsible for the rent price increase.

I started searching for data based on the premise that the displacement and the rent increase are the consequences of foreign people arriving in the city because they have the money to pay higher rents. But looking for information on Twitter, I realized that the arrival of foreigners is simply part or consequence of a larger problem instead of necessarily being the main reason. [...] Actually, it is a problem caused by real estate agencies that are not regulated. I found newspapers articles, and in one of the tweets, I found an article from 2017 by the (newspaper) Economist; it was very informative and described how this problem has been going on for a long time. And in this article, they mentioned the term the real estate cartel, and it was a very interesting term for me because I did not know about it, and it led me to find more information.- Lucas

The worksheets of the toolkit emphasize analyzing the data from Twitter to identify the multiple actors that might be involved in a social problem. Elisa's findings on real estate companies as relevant actors in the issue of gentrification led her to change the course of her initial examination. Therefore, instead of examining the role of foreigners in gentrification, she inquired about the government's role in controlling the increase in rent prices and searched for civil organizations that might collect information regarding the increment in the cost of housing and displacement of people in recent years.

8.2.3 Tensions on Finding Meaning on Twitter

Participants described the findings obtained through Twitter as significant starting points that revealed other actors and dimensions of the examined problem. Despite the value of data, participants reported being hesitant to draw conclusions from their findings, mainly due to the difficulty of tracking the speech of specific Twitter accounts over time, the potential for incomplete data, and the limited representation of data and its implications when using the findings obtained on Twitter. While some challenges are related to the design of the toolkit, which seeks to motivate a more qualitative analysis of the data, others are strictly related to the specific characteristics of Twitter. In response to these challenges, participants agreed that to make their findings from Twitter actionable and useful, they needed to conduct a more detailed analysis of the collected tweets, search for information in other sources, and supplement findings with interviews and other methodologies.

Tracking Changes Over Time

One of the strengths of Twitter that was of interest to practitioners is the possibility of obtaining data from the past to identify analogous cases of crises to understand how communities have solved problems in previous situations and inform current crises. While having access to tweets from the beginning of Twitter was perceived as a strength, one of the participants reported the difficulties of tracking over time the development of Twitter accounts founded in his initial exploration. As he explains in the following quote, his interest was to identify which accounts consolidated into non-profit organizations, transformed into citizen organizations, and vanished after the crisis.

> *But I think that the Toolkit and Twitter, in general, have many limitations that have to be carefully considered. The findings depend on the time window of the search and when the tweets were published. The frequency and content of tweets changed over time. In situations where there are many tweets, the changes in the conversation were very difficult to follow. It was difficult to follow multiple Twitter accounts and determine which ones were official, which ones were not, which ones transformed into a citizen organization, which ones continued over time, and which ones did not. So the findings on Twitter give starting points, not finishing points. So it's still a very worthwhile exploration, but it has many caveats.*- Lucas

Although the toolkit does not support conducting network analysis, nor does the manual include instructions or suggestions for making the type of analysis suggested by the participant, there are at least two alternatives that could be carried out with the current capabilities of the toolkit. A first option is to identify the Twitter accounts that want to be examined over time, download their tweet history, and then graph the history. Once the tweet history is generated, it would be possible to identify points in time where tweet activity intensified and then compare those moments with the timing of crises. Alternatively, tweets from the accounts of interest could be searched, using the time windows of different crises, and thus determine the behavior or reaction of the Twitter accounts concerning the different crises. This type of analysis could inform practitioners who could determine whether or not a Twitter account was consolidated in an organization depending on the type of activity recorded over time. Therefore, practitioners could determine whether or not a Twitter account was consolidated into an organization depending on the activity recorded over time.

The Latent Possibility of Having Incomplete Data

A second concern reported by the participants was the constant possibility of having an incomplete data set and the implications or consequences of not having all the information collected. One of the first clarifications included in the toolkit and then reinforced with the workshop discussion is that Twitter data is never complete due to several factors. As I mentioned before, despite how large the datasets of Twitter might be, the fact that not the entire population interacts on Twitter will always translate into having datasets with a limited representation of perspectives and needs. While participants were aware of Twitter constraints during the toolkit's evaluation, an additional concern emerged regarding not finding all the data on a specific topic on Twitter. Participants worried that not gathering and analyzing all the perspectives involved would diminish their analysis. Since Twitter's data searches are driven by keywords, Rolando expressed concern about not knowing which keywords he might be missing and the effect of that omission on the data sets.

> *Since the search works with keywords, there are many words I did not search,*

> *and therefore that was information I could not capture. For example, information on donation collection points is something that I did not find because, in my search, I used only the words: help, support, earthquake, Mexico. So, yes, Twitter partially works. And I think that this is also another thing to consider because I doubt that the 10 thousand tweets that I found about the earthquake using those three search words have given me the complete universe of data on the earthquake and the citizen response. In other words, I missed several tweets that did not contain those keywords.*- Rolando

Although Rolando acknowledges that the tweets collected in the initial search hardly reflect the complete universe of tweets regarding the citizen response to the earthquake, his comment also reflects the worrisome of missing information since he did not include the keywords of donation collection points in his initial search. Ultimately, what this quote shows is that there is an interest in obtaining data that is complete. Otherwise, the incompleteness of data could affect the reliability of the conclusions that practitioners could reach.

> *I feel I did a good exploration, but I don't know what's true and what's not about the conclusions I come to with the analysis. How accurate or not is the reality of the tweets that are in the universe of Tweets? For example, the finding that millennials were an important demographic group to help in collective organizing is relevant to me because I saw tweets about it. However, how can I be sure that the role of this group of people was so impactful to the collective organization in the city and that without the Millenials, the response would have been completely different? So how can I determine the scope of the results or how to be sure that Twitter did not give me a false idea of the phenomenon?*- Rolando

Since it has been established that the data obtained from Twitter will never be complete, it is worth rethinking what constitutes having a complete data set. From practitioners, the concern of having complete data is linked to obtaining findings from Twitter that are more precise and actionable. In the context of this research, where the objective is to use the data from Twitter to inform the work of NGOs, one could think of a notion of completeness of data that considers both the limitations of data coming from Twitter and the interests of the practitioners. Therefore, instead of defining complete data in the strict sense of including the entire universe of tweets produced on a topic during a window of time, one could think of a definition of completeness that refers to having represented the different communities involved and the perspectives of various actors.

In this sense, the toolkit could encourage practitioners to think of Twitter data as a dataset that is always iterative, constantly being built, and with each iteration, perspectives and content from different stakeholders are added. Thereby motivating practitioners not to hold the Twitter data set with the same standards that a census or other traditional data set would have. But instead, think in a different sense of completeness. The following comment from one of the participants echoes this approach of completeness.

> *The information that we can extract from Twitter is small signs or glimpses of stories that are a little bit more complex. I mean, it's not just the information that tweets can contain links or hashtags. Their content also speaks a lot about relationships, grudges, revenge, anguish, and frustration, especially in the way that Twitter is used as a complaint mechanism or to make visible communities not served or made invisible by others. So, in that sense, I find it very interesting to see Twitter as a telescope for that collective social dimension and I try to situate where we are regarding social problems that are either sensitive or have not been addressed. Problems are not equitably represented by all their stakeholders on Twitter. Therefore, more than having concrete information on people's perspectives about an issue, Twitter helps us to put points on the map, and I suppose that each topic will be different depending on how it manifests*

on Twitter.- Avellaneda

8.2.4 Final Remarks

Based on these findings, we can conclude that identifying relevant content requires practitioners to determine the appropriate keywords or analogous cases that describe the problem according to how people discuss it on Twitter. However, this requires participants to spend considerable time exploring and using different strategies to understand how their target communities interact on Twitter.

The second set of lessons relates to the value of information from Twitter. Participants described the findings obtained through Twitter as significant starting points that revealed other actors and dimensions of the examined problem and evidence of the differences in how communities react to crises.

Despite the value of data, participants reported being hesitant about the accuracy of their findings, mainly due to the potential for incomplete data and the limited representation of data and its implications when using the results obtained on Twitter. In response to these challenges, participants agreed that to make the findings from Twitter actionable and useful, they needed to conduct a more detailed analysis of the collected tweets, search for information in other sources and supplement their findings with interviews and other methods.

CHAPTER 9
CONCLUSIONS

The outcome of my research provides an initial understanding of the implications of mobilizing social media data to inform the work of non-profit and civil organizations. Building on the insights obtained in my fieldwork, I proposed a mixed methodology and a set of tools that interwove qualitative and computational techniques to make sense of data from Twitter and use it in institutional contexts. Both the toolkit and the methodology encourage practitioners to engage with local perspectives and conditions when making sense of data [19]. In what follows, I conclude my dissertation by reflecting on some of the tensions I observed during the toolkit's evaluation. Then, I provide a set of guidelines that aim to provide an overview of the topics and problems that are more suitable to be examined using data from Twitter, and I conclude this section with potential next steps.

9.0.1 Maintaining the Locality of Data is a Work of Care

The design of the toolkit and the development of the methodology described in this work responded to the call of maintaining the locality of data by promoting a situated and critical analysis that recognizes the role of human intervention in making sense of data from Twitter. During the workshops and interviews, it became more evident that as participants spent more time with the toolkit, exploring their topics of interest on Twitter, reframing their inquiries, and making sense of the data, they became more critical of the limitations of using this type of data. As participants became more aware of the implications of using data from Twitter, various tensions emerged that challenged the actual value that this data might have for practitioners.

One of the most prominent tensions was that conducting a qualitative analysis demands an extraordinary amount of time, care, and labor. Among the central arguments that I have made along this research is that to use data from any social media platform to inform other actors or stakeholders, it is imperative to first understand how topics are discussed and then how that fits with the definition of evidence of the organizations that might be able to leverage such content. Although I consider this to be a crucial step in making sense of social media data, achieving this requires conducting multiple and iterative analyses, demanding time, and care from the practitioners to reach an agreement on the definition of evidence. Even though the toolkit included a tool that enabled filtering the tweets and identifying the most similar to a specific topic (Module 2), for some of the participants, the amount of time required was still significant.

Additionally, participants expressed concern about spending too much time doing initial explorations and that the result would not answer the specific questions they were seeking to answer. Participants recognize that any finding from Twitter should be considered as starting point that needs to be completed with additional information using other sources of data and methods. In conclusion, although participants found the obtained insights from Twitter interesting, the fact that, more time and work was required to transform those findings into actionable information represented a constant challenge.

9.0.2 Guidelines For Practitioners

Deciding whether it is appropriate to use data from Twitter or other social platforms depends on 1) the problem practitioners aim to examine, 2) how they plan to use the data, and 3) the conditions of the examined communities.

Recommendations to Determine Appropriate Problems to Examine Using Data from Twitter

While it is possible to search any number of keywords to find out about different topics on Twitter, based on the insights gathered during the fieldwork described in this dissertation, we can conclude that the problems with the following characteristics are more suitable to be examined using data from Twitter.

1. Ongoing local crises that affect the bulk of the population

2. Growing situations in which citizens request the government or any other authority to make changes because they are dissatisfied with social or economic conditions and political decisions.

3. Natural disasters, human-made crises such as wars and terrorist attacks and chemical and viruses disease crises: As was previously mentioned, the research on Crisis Informatics provides evidence that in the aftermath of natural disasters, social media platforms have become a routine part of crisis response and a formalized infrastructure for various tasks such as collecting witness accounts, monitoring the media, and providing updates during crisis responses.

4. Crises related to human rights violations.

Problems that are not suitable to be examined with data from Twitter are those considered systemic social problems such as poverty, corruption, or unemployment. I recommend reframing from analyzing these types of topics because the discourse and data on Twitter mainly reflect trends and ongoing situations rather than normalized practices in a specific context. Moreover, these types of problems are entrenched, complex, and the result of multiple issues rather than a single factor, which further complicates its examination using Twitter data.

The second category of problems that is not advisable to examine using Twitter data is those concerning a specific population group filtered by gender, age group, or profession. Due to the nature of the platform, it is difficult to determine characteristics such as profession, gender, or age group of users from the data they share on Twitter. For example, if the interest is to understand the perspective of teenage girls (13 to 19 years old) or people with a specific profession on any topic, then Twitter will not be a recommended source.

Suggestions on Defining how to Use the Data from Twitter

A second aspect to consider when defining what topics are the most appropriate to examine using data from Twitter relates to how practitioners plan to use the findings from their explorations on Twitter. Practitioners ought to reflect on the type of arguments they aim to make and the evidence they might need to support them. In this sense, it is imperative to distinguish between using the evidence from Twitter to inform other stakeholders (e.g., police and other non-profit organizations) or their team to define the next steps. Regardless of the definition of evidence, Twitter can be an appropriate source of insights in case practitioners aim to understand and collect data on any communities' aspects that fall into the following categories:

1. People's experiences and strategies during crises: Understand with greater detailed how communities experience, react and solve local problems.

2. Actors involved: Identify actors and stakeholders that contribute to the resolution of local crises that were not previously identified.

3. Discourse: Learn about the multiple opinions of an ongoing or past crisis, and identify communities' perspectives that are not often considered, e.g., indigenous groups.

In contrast, it is not recommended to use Twitter data if the ultimate purpose is any of the following:

1. Quantify social, cultural, economic, or political conditions of a population: Twitter should not be used as a sole source of information to quantify social phenomena. As explained before, due to the format of Twitter data, it is always incomplete and never exhaustive of people's experiences.

2. Answer exact concerns of a determined population: As mentioned before, solely using data from Twitter is not recommended to gather data from communities classified by age, gender, profession, or location.

Considerations on the Conditions of Communities

A final set of considerations to be made when deciding the topics to analyze using Twitter is related to the conditions of the examined communities regarding their quality of Internet access, their level of smartphone ownership, and their preferences of social media platforms.

Considering these three aspects is essential since they determine who accesses Twitter, their frequency, and even their skills to organize using digital technologies. Although meeting these conditions is highly desirable, if they are not met, practitioners could still consider examining data from Twitter to inform their work; as long as they take the necessary measures to counteract the potential biases that the limited access to the internet and smartphones might impose on the studied communities.

9.0.3 Next Steps

The outcome of this dissertation has set the stage to continue exploring what it means to design tools that advocate for maintaining the locality of data. The evaluation and findings presented in the previous chapter represent only the initial understanding of what it entails to integrate the content from Twitter into the institutional context of organizations. As I mentioned in Chapter 8, the evaluation of the toolkit with the non-profit organizations CIEP and Social TIC will continue. In partnership with the Accelerator Lab, we outlined a work plan to continue testing the toolkit and evaluate the implications of having the Accelerator lab adopt and promote the toolkit. As I continue the collaboration with the Accelerator Lab, I expect to develop a more detailed understanding of the limitations of the toolkit and the overarching constraints of using social media data to inform non-profit organizations.

In conclusion, based on the evaluation lessons, the potential next steps are:

1. To include the strategies that participants developed into the toolkit and evaluate them.

2. Conduct more evaluations of the toolkit with different types of organizations that focus on diverse topics.

3. Based on the results of evaluations, develop a taxonomy of topics that are more appropriate to be informed using data from Twitter.

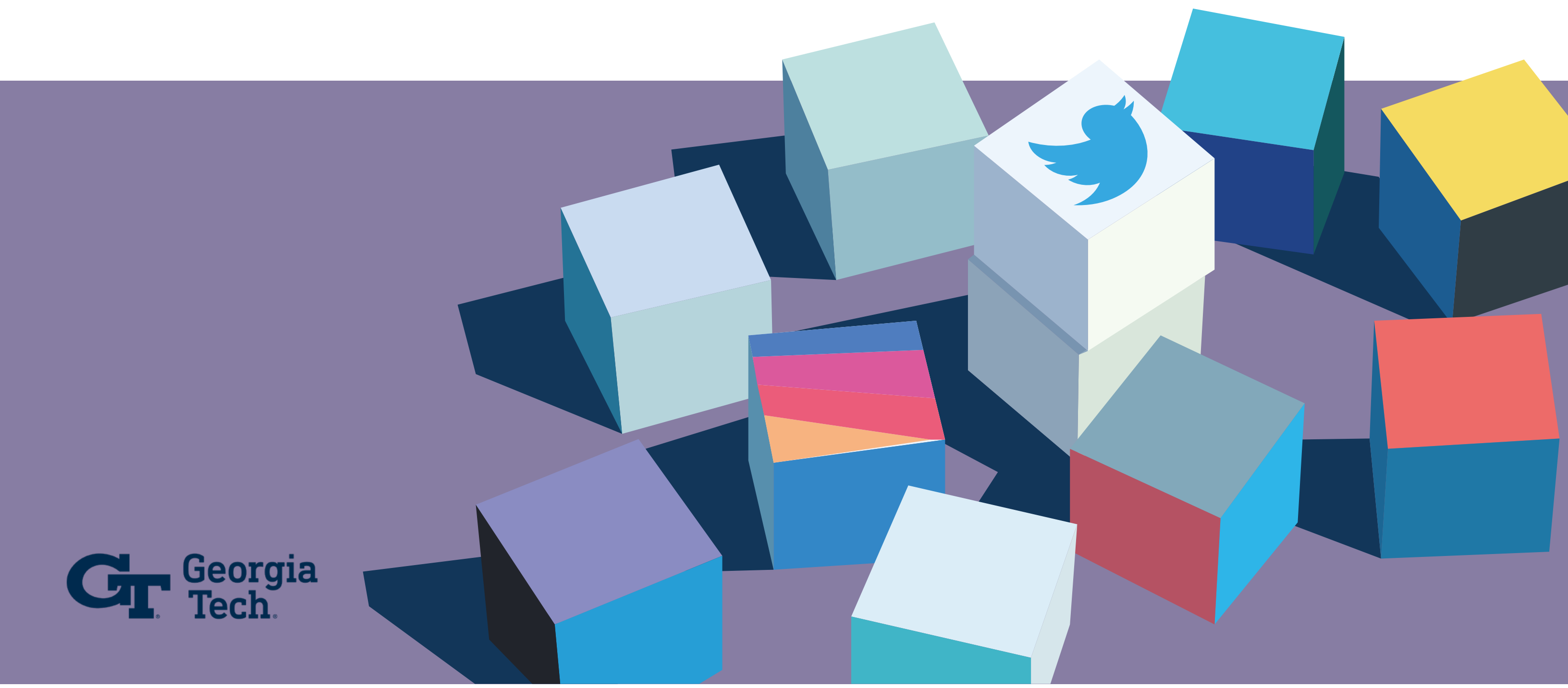
Appendices

APPENDIX A

MANUAL BITÁCORA TOOLKIT

Manual

bitácora Toolkit



Semblanza

Adriana Alvarado Garcia es doctora en **Computación Centrada en el Humano** (Georgia Tech). Su investigación doctoral se ha enfocado en examinar las motivaciones de las comunidades para utilizar las plataformas de redes sociales como un medio de organización y las implicaciones de conectar los datos y experiencias recolectadas en las plataformas con instituciones que podrían aprovecharlas como insumos para informar su trabajo con comunidades.

En el 2019 fue nombrada Data Fellow del primer programa organizado por la Organización de las Naciones Unidas en colaboración con Global Pulse. A partir de enero del 2020 colaboró con el Laboratorio de Aceleración de la Ciudad de México que forma parte del Programa de las Naciones Unidas para el Desarrollo (PNUD).

Estas colaboraciones le permitieron examinar el rol de las prácticas organizativas y la interpretación humana cuando las organizaciones no gubernamentales (ONG) utilizan datos de redes sociales para informar sus intervenciones. Aprovechando las lecciones de estas colaboraciones, Adriana desarrolló el kit de herramientas Bitacora, que tiene por objetivo apoyar a las ONG para que utilicen los rastros digitales como evidencia para descubrir la pluralidad de formas de saber y hacer de las ciudadanas en tiempos de crisis centrándose en las perspectivas de la comunidad y el contexto de la producción de datos.

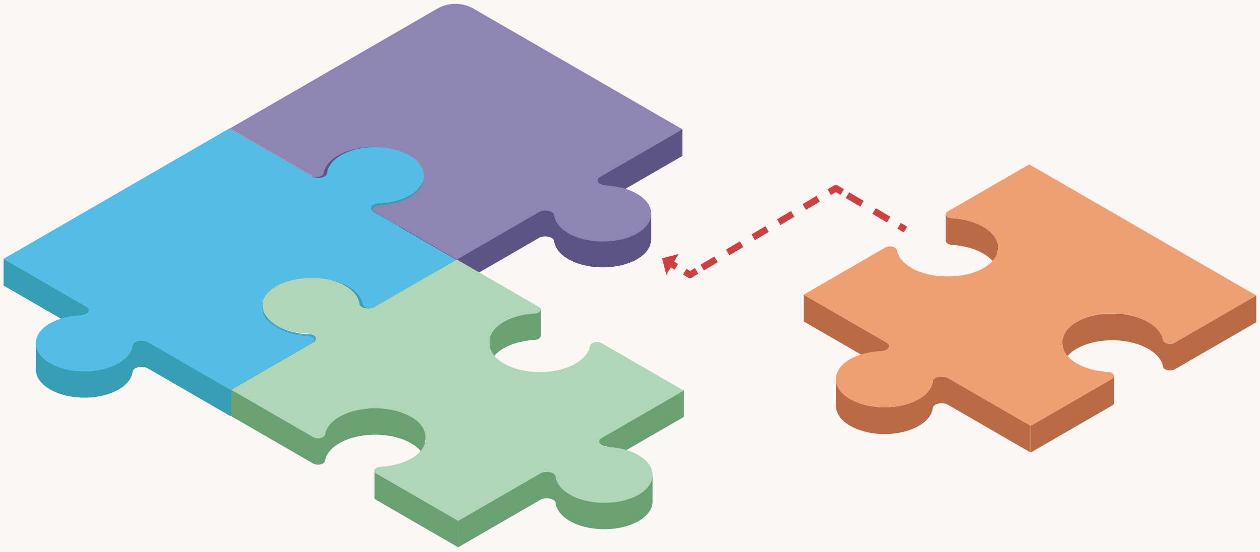
Datos de contacto
Correo electrónico: **alvaradogarcia.adriana@gmail.com**
Twitter: **@ella_es_adriana**

Créditos

Este kit de herramientas fue diseñado como parte de la investigación doctoral titulada Diseño de tecnologías centradas en el ser humano para migraciones de datos de redes sociales por **Adriana Alvarado García**.

Diagramación por Katy Garfias.
Correo electrónico: katygarfiasg@gmail.com

01 INTRODUCCIÓN AL MANUAL

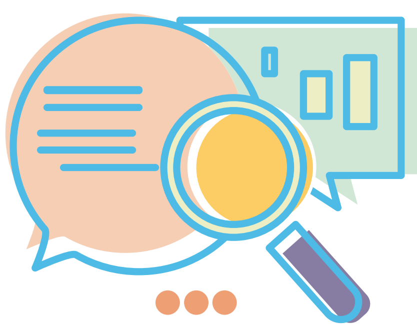


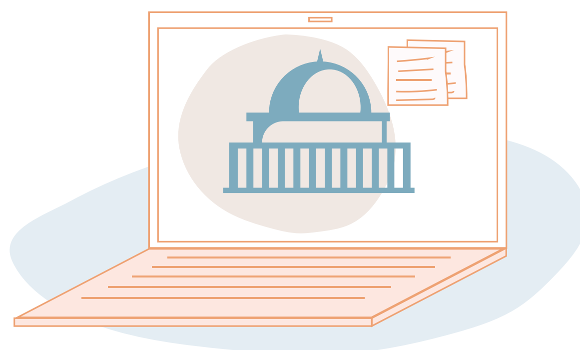
Introducción a las herramientas

¿Que es Bitácora?

Bitácora es un kit de herramientas computacionales y cualitativas que tienen por objetivo facilitar el uso de datos de redes sociales para informar el trabajo de las instituciones gubernamentales, organizaciones civiles y organizaciones no gubernamentales (ONGs).

El propósito del kit es guiar a las lectoras en el proceso de distinguir las perspectivas de la comunidad y el contexto de la producción de datos, al momento de recolectar e interpretar datos de redes sociales, específicamente en Twitter.





¿Cuáles son los componentes de bitácora?

Los componentes del kit son:

1.- Una plataforma que contiene dos herramientas computacionales.

Módulo 1: Herramienta para buscar y descargar tuits.

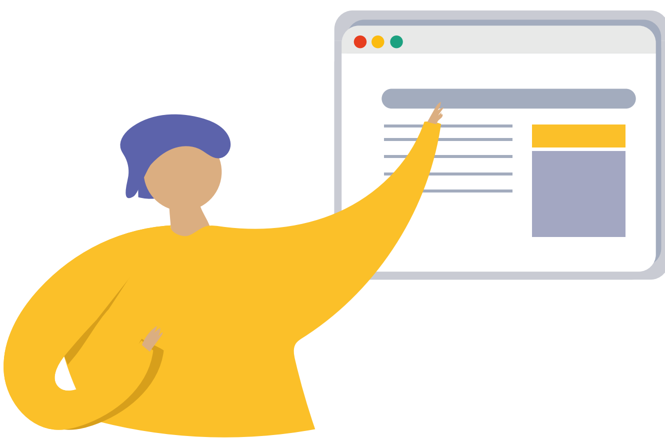
Módulo 2: Herramienta que permite el análisis de tuits a escala utilizando procesamiento de lenguaje natural. Esta herramienta facilita el filtrado e identificación de tuits con contenido similar al problema que se está examinando.

2.- Este manual que concentra una serie de aprendizajes que guían a los lectores en el proceso de utilizar las herramientas de la plataforma para buscar, recolectar, interpretar e integrar datos de redes sociales como evidencia para informar su trabajo.

Introducción a las herramientas

¿Para quién está diseñado el kit de herramientas?

Creamos este conjunto de herramientas para profesionales que trabajan en organizaciones de la sociedad civil, organizaciones sin fines de lucro e instituciones gubernamentales que estén interesadas en aprovechar el potencial de las redes sociales para identificar capacidades locales, monitorear crisis comunitarias y desarrollar intervenciones basadas en prácticas locales.



¿Para qué es este kit de herramientas?

Si trabajas en problemas...

• Que son difíciles de examinar utilizando fuentes de datos tradicionales

• Para los que no existen datos actualizados

Si te interesa...

• Entender la respuesta de la comunidad a problemas locales

• Entender las conexiones y relaciones entre individuos que son difíciles de rastrear a través de datos estadísticos

• Entender el discurso sobre un tema específico

• Identificar personas, capacidades, narrativas y modelos locales de organización

• Identificar actores y acciones que no estaban previamente identificados

• Reconocer la riqueza de enfoques y la pluralidad de formas de conocer y hacer de comunidades en tiempos de crisis

• Diferenciar cómo se manifiesta un mismo fenómeno en distintas comunidades

Introducción a las herramientas

¿Cómo utilizar este manual?

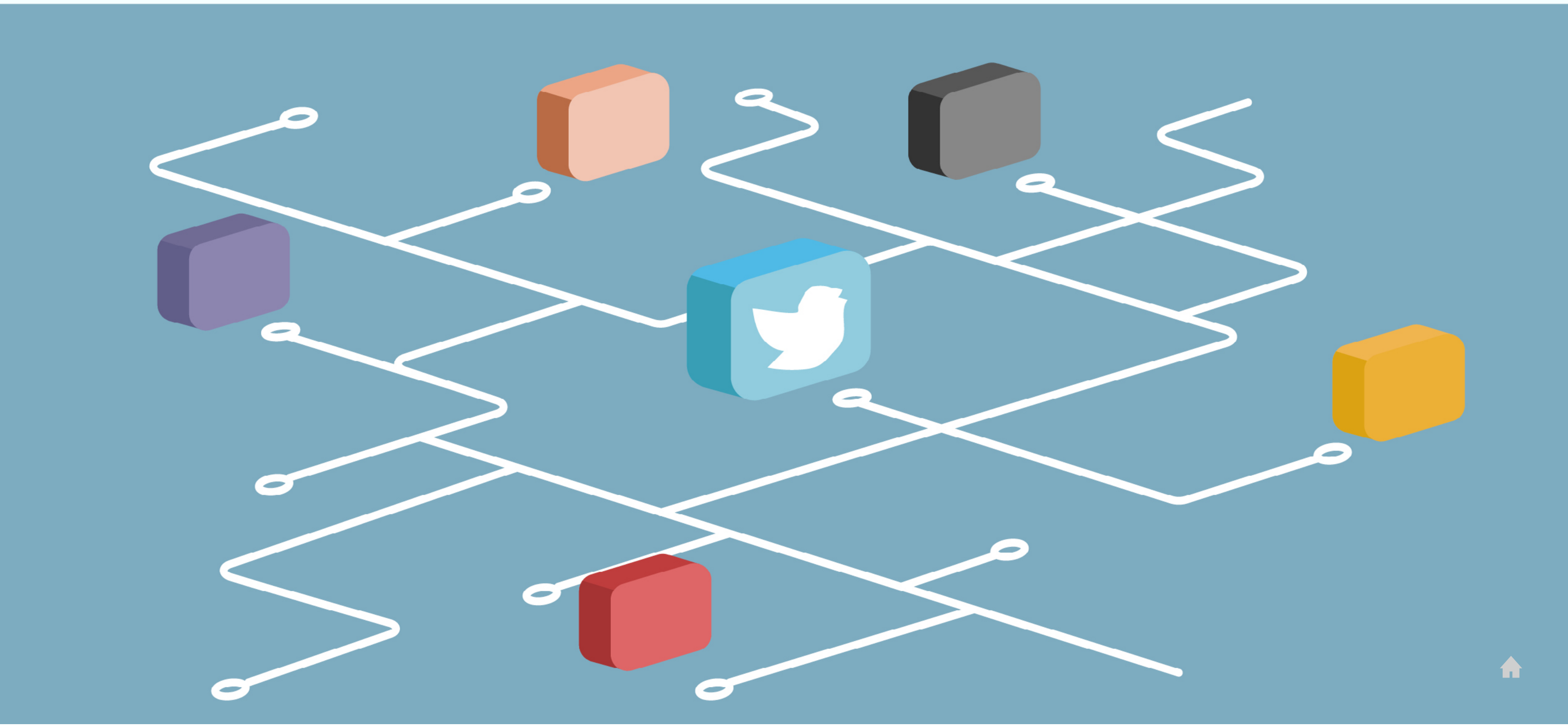
El propósito de este manual es brindar una introducción y una guía a las y los lectores tanto del uso de herramientas de bitácora así como de las ventajas y limitaciones que supone el uso de datos provenientes de redes sociales. El manual incluye una serie de recursos que promueven un uso crítico y contextualizado de los datos de Twitter.

El manual esta dividido en cuatro secciones:

Se recomienda leer la sección 1 y 2 del manual antes de interactuar con las herramientas de la plataforma, ya que es necesario tener un entendimiento de lo que es o no posible hacer con los datos de Twitter antes de buscar, descargar y analizar datos.

En el manual se incluyen también varias plantillas de trabajo que se recomiendan utilizar en paralelo con las herramientas computacionales. Sugerimos no solo utilizar las plantillas sino también adaptarlas a las particularidades y necesidades de tu investigación.

02 INTRODUCCIÓN A LAS REDES SOCIALES



Entendiendo el potencial de las redes sociales

¿Cómo pueden las redes sociales ayudar a informar el trabajo de las instituciones gubernamentales, ONGs y organizaciones civiles?

En los últimos años, las plataformas de redes sociales han demostrado ser una herramienta efectiva que permite a las comunidades desarrollar nuevas formas de organización mediante el intercambio y acumulación de datos. Debido a sus características particulares, el contenido generado en redes sociales se ha convertido en una alternativa para informar las intervenciones de organizaciones no gubernamentales, civiles y de gobierno. Los datos de redes sociales tienen el potencial de informar sobre el contexto de las necesidades y problemáticas de poblaciones, así como sobre las capacidades y la experiencia de los ciudadanos al responder a crisis locales.

En este manual nos enfocamos únicamente en los datos de Twitter, los cuales permiten identificar: personas, activos locales, narrativas, ritmos y modelos de organización que de otra manera son sido difíciles de encontrar.

El análisis de los **datos de Twitter** ofrece una riqueza única de enfoques y pluralidad de formas de conocer y hacer de comunidades en tiempos de crisis.

Entendiendo el potencial de las redes sociales

A continuación describimos algunos proyectos que han utilizado datos de Twitter como principal fuente de información y que ilustran las principales fortalezas de Twitter.

Caso de estudio:

Uso de Twitter para medir la discusión global sobre el cambio climático

Descripción: Global Pulse desarrolló un monitor de redes sociales en tiempo real para medir y explorar el discurso en Twitter sobre el cambio climático en apoyo de la Cumbre del Clima de las Naciones Unidas en septiembre de 2014. La medición y visualización de los tweets públicos a lo largo del tiempo creó una línea de base de participación y mostró un aumento significativo en debates sobre el cambio climático durante la Cumbre del Clima.

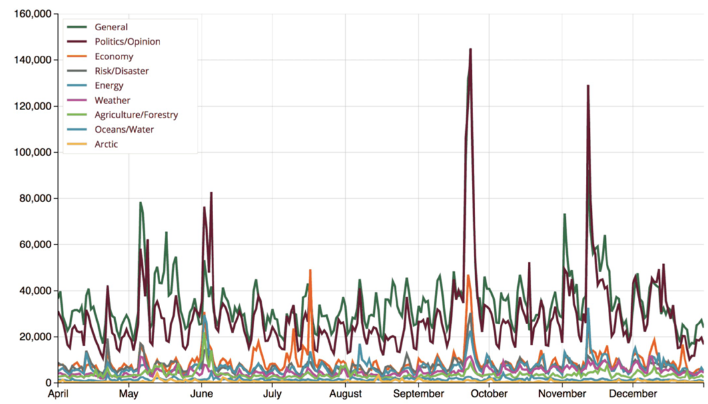
Año: 2014

Autores: UN Global Pulse

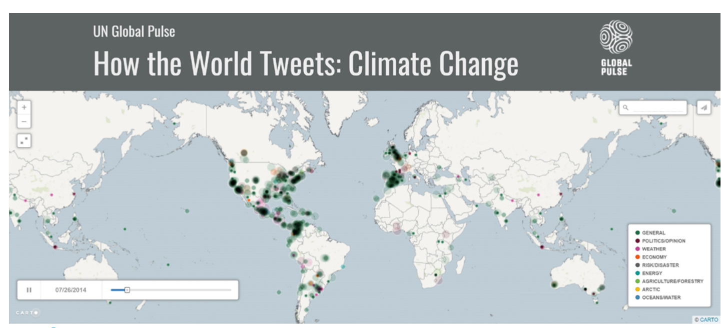
Referencia: UN Global Pulse, 'Using Twitter to Measure Global Engagement on Climate Change', Global Pulse Project Series, no.7, 2015.

Link: https://www.unglobalpulse.org/wp-content/uploads/2014/09/UNGP_ProjectSeries_Climate_Monitor_2015_0.pdf

Aprendizaje: Los datos de Twitter pueden ayudar a medir la conciencia, respaldar la toma de decisiones sobre políticas de distinta índole e impulsar una mayor participación pública.



Gráfica de los volúmenes diarios de tweets en inglés sobre el cambio climático y varios subtemas relacionados.



Monitor de redes sociales sobre cambio climático.

Entendiendo el potencial de las redes sociales

Caso de estudio:

Uso de datos de Twitter para analizar el sentimiento público sobre la reforma de la política de subsidios a los combustibles en El Salvador

Descripción: En 2011, El Salvador hizo reformas de política a un subsidio nacional al gas propano, lo que provocó un descontento público generalizado y una serie de huelgas por parte de las empresas distribuidoras. El Banco Mundial y Global Pulse colaboraron en un proyecto de investigación que analiza el contenido y el sentimiento de los tweets para comprender mejor la opinión pública en torno a las reformas.

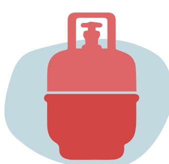
Año: 2015

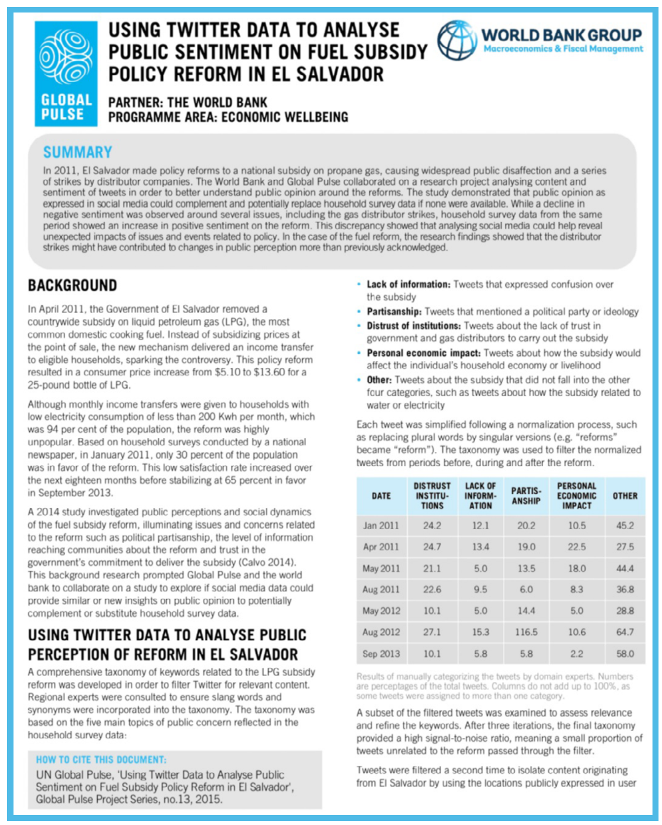
Autores: UN Global Pulse New York

Referencia: UN Global Pulse, 'Using Twitter Data to Analyse Public Sentiment on Fuel Subsidy Policy Reform in El Salvador', Global Pulse Project Series, no.13, 2015.

Link: https://www.unglobalpulse.org/wp-content/uploads/2015/02/using-twitter-data-to-analyse-public-sentiment-on-fuel-subsidy-policy-reform-in-el-salvador.pdf

Aprendizaje: El análisis de las redes sociales podría ayudar a revelar impactos inesperados de problemas y eventos relacionados con la política.





USING TWITTER DATA TO ANALYSE PUBLIC SENTIMENT ON FUEL SUBSIDY POLICY REFORM IN EL SALVADOR

GLOBAL PULSE

WORLD BANK GROUP
Macroeconomics & Fiscal Management

PARTNER: THE WORLD BANK
PROGRAMME AREA: ECONOMIC WELLBEING

SUMMARY

In 2011, El Salvador made policy reforms to a national subsidy on propane gas, causing widespread public disaffection and a series of strikes by distributor companies. The World Bank and Global Pulse collaborated on a research project analysing content and sentiment of tweets in order to better understand public opinion around the reforms. The study demonstrated that public opinion as expressed in social media could complement and potentially replace household survey data if none were available. While a decline in negative sentiment was observed around several issues, including the gas distributor strikes, household survey data from the same period showed an increase in positive sentiment on the reform. This discrepancy showed that analysing social media could help reveal unexpected impacts of issues and events related to policy. In the case of the fuel reform, the research findings showed that the distributor strikes might have contributed to changes in public perception more than previously acknowledged.

BACKGROUND

In April 2011, the Government of El Salvador removed a countrywide subsidy on liquid petroleum gas (LPG), the most common domestic cooking fuel. Instead of subsidizing prices at the point of sale, the new mechanism delivered an income transfer to eligible households, sparking the controversy. This policy reform resulted in a consumer price increase from $5.10 to $13.60 for a 25-pound bottle of LPG.

Although monthly income transfers were given to households with low electricity consumption of less than 200 Kwh per month, which was 94 per cent of the population, the reform was highly unpopular. Based on household surveys conducted by a national newspaper, in January 2011, only 30 percent of the population was in favor of the reform. This low satisfaction rate increased over the next eighteen months before stabilizing at 65 percent in favor in September 2013.

A 2014 study investigated public perceptions and social dynamics of the fuel subsidy reform, illuminating issues and concerns related to the reform such as political partisanship, the level of information reaching communities about the reform and trust in the government's commitment to deliver the subsidy (Calvo 2014). This background research prompted Global Pulse and the world bank to collaborate on a study to explore if social media data could provide similar or new insights on public opinion to potentially complement or substitute household survey data.

USING TWITTER DATA TO ANALYSE PUBLIC PERCEPTION OF REFORM IN EL SALVADOR

A comprehensive taxonomy of keywords related to the LPG subsidy reform was developed in order to filter Twitter for relevant content. Regional experts were consulted to ensure slang words and synonyms were incorporated into the taxonomy. The taxonomy was based on the five main topics of public concern reflected in the household survey data:

HOW TO CITE THIS DOCUMENT:

UN Global Pulse, 'Using Twitter Data to Analyse Public Sentiment on Fuel Subsidy Policy Reform in El Salvador', Global Pulse Project Series, no.13, 2015.

- **Lack of information:** Tweets that expressed confusion over the subsidy
- **Partisanship:** Tweets that mentioned a political party or ideology
- **Distrust of institutions:** Tweets about the lack of trust in government and gas distributors to carry out the subsidy
- **Personal economic impact:** Tweets about how the subsidy would affect the individual's household economy or livelihood
- **Other:** Tweets about the subsidy that did not fall into the other fcur categories, such as tweets about how the subsidy related to water or electricity

Each tweet was simplified following a normalization process, such as replacing plural words by singular versions (e.g. "reforms" became "reform"). The taxonomy was used to filter the normalized tweets from periods before, during and after the reform.

DATE	DISTRUST INSTITU-TIONS	LACK OF INFORM-ATION	PARTIS-ANSHIP	PERSONAL ECONOMIC IMPACT	OTHER
Jan 2011	24.2	12.1	20.2	10.5	45.2
Apr 2011	24.7	13.4	19.0	22.5	27.5
May 2011	21.1	5.0	13.5	18.0	44.4
Aug 2011	22.6	9.5	6.0	8.3	36.8
May 2012	10.1	5.0	14.4	5.0	28.8
Aug 2012	27.1	15.3	116.5	10.6	64.7
Sep 2013	10.1	5.8	5.8	2.2	58.0

Results of manually categorizing the tweets by domain experts. Numbers are perceptages of the total tweets. Columns do not add up to 100%, as some tweets were assigned to more than one category.

A subset of the filtered tweets was examined to assess relevance and refine the keywords. After three iterations, the final taxonomy provided a high signal-to-noise ratio, meaning a small proportion of tweets unrelated to the reform passed through the filter.

Tweets were filtered a second time to isolate content originating from El Salvador by using the locations publicly expressed in user

Poster del proyecto

Entendiendo el potencial de las redes sociales

Caso de estudio:

Explorando el papel de los lazos sociales en la reacción de una comunidad ante la pandemia de COVID-19 a través de Twitter

Descripción: A medida que se desarrollaba la pandemia por el COVID-19, no existían fuentes oficiales de datos que retrataran las necesidades y estrategias de los ciudadanos para afrontar e innovar los desafíos impuestos por la crisis. En esta investigación, se analizaron más de 700 tuits como una fuente alternativa para capturar distribución geográfica e impacto de iniciativas y estrategias ciudadanas inusuales que surgieron en respuesta a las afectaciones de la pandemia.

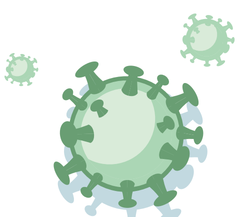
Año: 2020

Autores: Laboratorio de Aceleración de México

Referencia: Alvarado, A., & Munguía, J. (2021, November 25). Exploring the role of social ties in a community's reaction to the COVID-19 pandemic through Twitter. UNDP AccLab Mexico.

Link: https://www.mx.undp.org/content/mexico/es/home/blog/20210/11/exploring-the-role-of-social-ties-in-a-communitys-reaction-to-th.html

Aprendizaje: Esta exploración inicial puso de manifiesto la riqueza de enfoques y la pluralidad de formas de saber y hacer en tiempos de crisis.



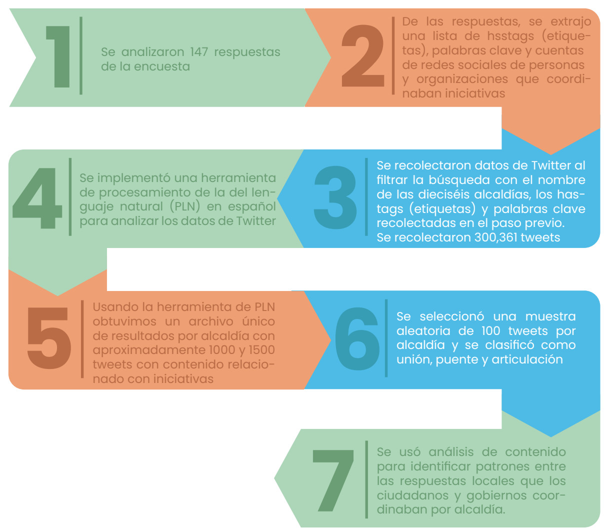


Diagrama que muestra el proceso seguido para la exploración de Twitter.

Entendiendo el potencial de las redes sociales

Caso de estudio:

Explorando el uso de datos de redes sociales como evidencia para la defensa de los derechos humanos

Descripción: Esta investigación consistió en utilizar los reportes ciudadanos de personas desaparecidas distribuidos en grupos de Facebook como una fuente alternativa para completar registros oficiales sobre la actual crisis de desapariciones en México.

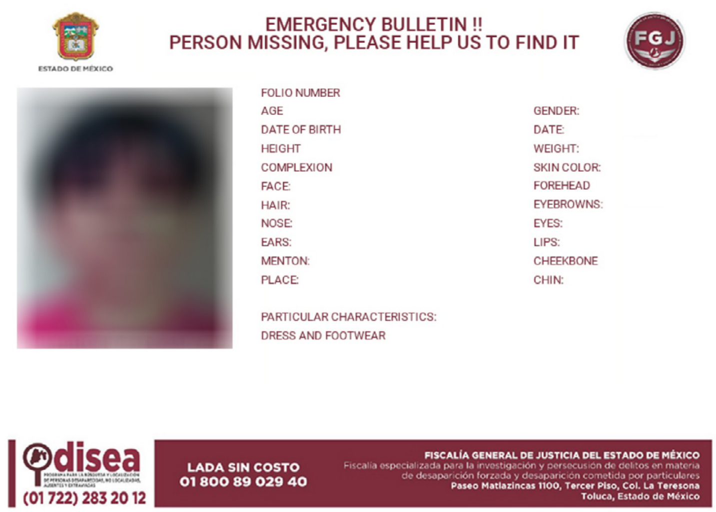
Año: 2020

Autores: Adriana Alvarado Garcia, Matthew J. Britton, Dhairya Manish Doshi, Munmun De Choudhury, and Christopher A. Le Dantec.

Referencia: 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. Interaction 4, CSCW3, Article 268 (December 2020), 25 pages.

Link: https://dl.acm.org/doi/10.1145/3434177

Aprendizaje: Esta investigación abona a una mejor comprensión de los desafíos y oportunidades de utilizar el conocimiento local de las comunidades en línea para ser utilizado como evidencia en contextos de violaciones de derechos humanos.



Ejemplo de un volante de una persona desaparecida emitido por el gobierno de México (Alvarado, 2021)

¿Cómo determinar si los datos de Twitter son los adecuados para informar tu trabajo?

Utiliza el siguiente diagrama de árbol de decisión para decidir si los datos de Tweet son o no la fuente de información más adecuada para informar tu trabajo.

Inicio → Planeo utilizar los datos de Twitter para...

→ Cuantificar aspectos sociales, culturales, económicos o políticos de una población en particular. → Debido al formato de los datos de Twitter, estos siempre están incompletos y nunca son exhaustivos de las experiencias de los ciudadanos. Es por lo que no deben utilizarse como fuente para cuantificar un fenómeno social.

↓

- Recopilar narrativas
- Entender con mayor granularidad cómo las comunidades experimentan los retos o crisis locales que enfrentan.
- Identificar actores y acciones que no estaban previamente identificados, es decir como fuente alternativa de datos y narrativas.
- Mapear datos que aún no existen en fuentes oficiales
- Identificar estrategias colectivas para afrontar problemáticas locales.

↓

Usando los datos de Twitter quiero entender...

→ Un problema sistémico como: pobreza, corrupción, desempleo, etc. → Twitter no es una fuente recomendable para examinar problemas sistémicos, ya que los datos de esta plataforma reflejan sobre todo temas destacados o tendencias por lo que resulta mas complejo examinar problemas sistémicos o prácticas que se han normalizado en un contexto en específico. Para examinar problemas sistémicos considera usar otras fuentes de datos.

↓

El discurso de un evento que sucedió o que está sucediendo

↓

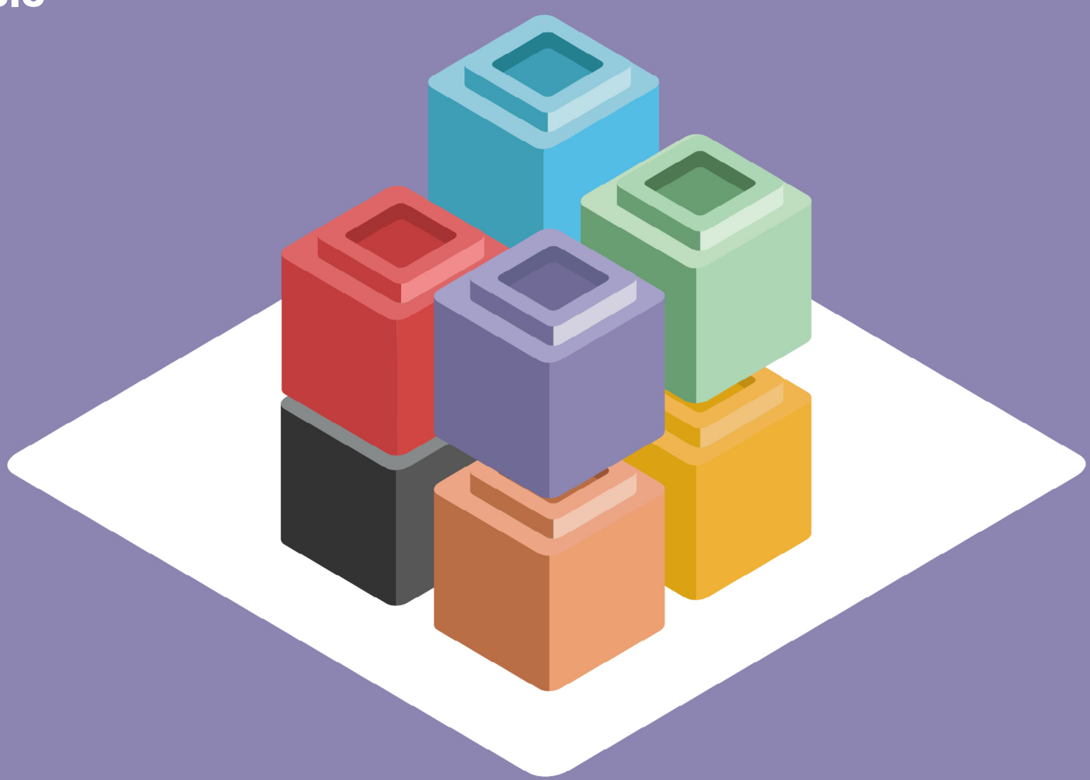
Me interesa entender la perspectiva de...

→ Poblaciones especificas filtradas por género, grupo de edad, o profesión. Por ejemplo, me interesa entender la perspectiva de adolescentes en temas de violencia de genero. → Debido a la naturaleza de Twitter resulta imposible determinar características específicas, como son profesión, género o grupo de edad, de los usuarios que generan datos en la plataforma.

↓

Distintas comunidades o los diversos actores involucrados en el problema que estoy examinando → Con base a tus respuestas, los datos de Twitter pueden ser una fuente alternativa de información para informar tu trabajo.

03 ANÁLISIS



Análisis

Introducción

Los datos provenientes de las redes sociales son siempre conjuntos de datos incompletos, ya que debido a su naturaleza no son completamente representativos sino que reflejan únicamente las perspectivas, opiniones y necesidades de una limitada muestra demográfica, específicamente de quienes usan redes sociales. Hay además siempre una restricción impuesta por las propias plataformas de redes sociales quienes determinan el número de datos que es posible recolectar. A pesar de estas limitaciones, los datos de Twitter pueden proporcionar evidencia de las respuestas colectivas, capacidades y particularidades de las comunidades que de otro modo, por ejemplo usando fuentes más tradicionales como las estadísticas gubernamentales, serían difíciles de identificar.

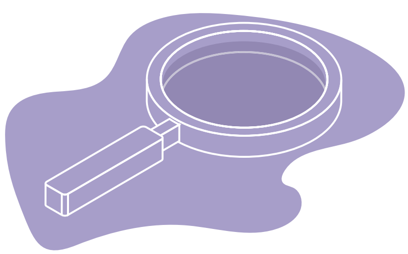
En esta sección se describe el proceso para conducir un análisis que:

1. Distinga las perspectivas de la comunidad y considere el contexto de la producción de los datos

2. Reconozca las limitaciones de los datos de redes sociales en el proceso de interpretación

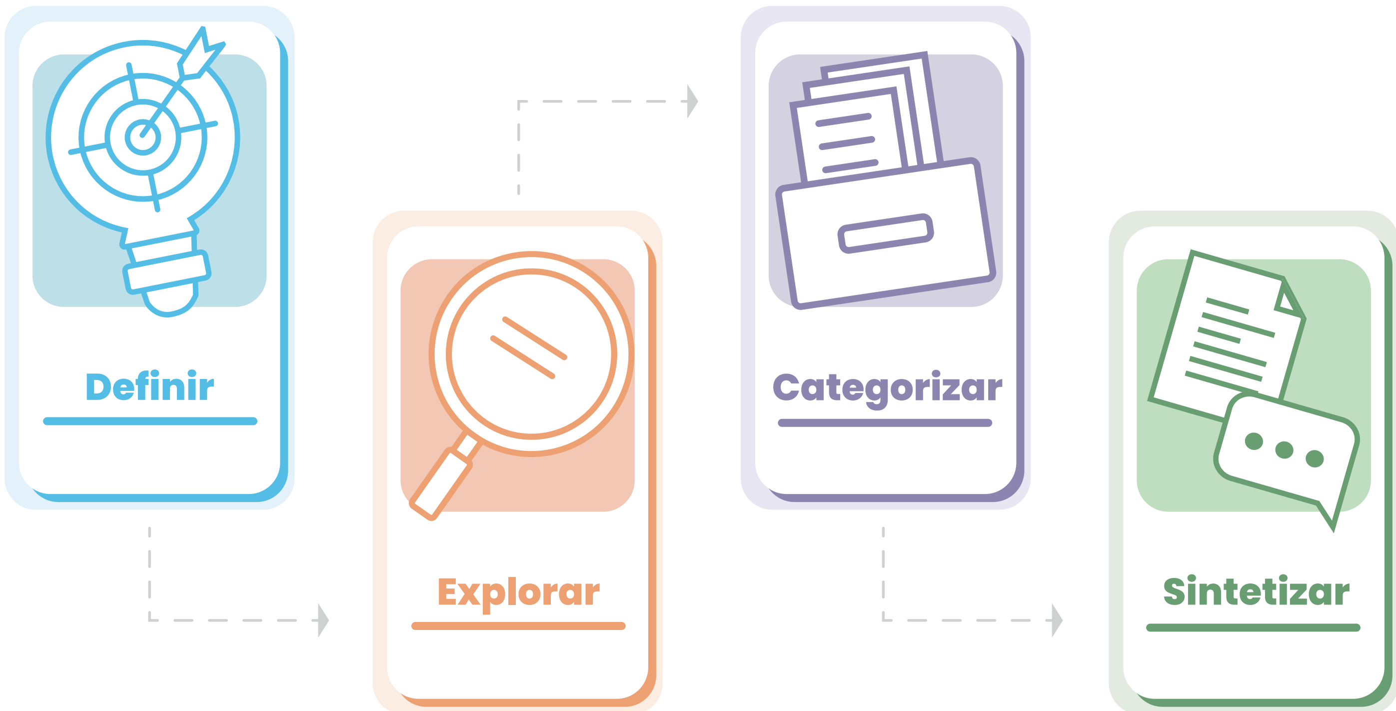
Esta sección se divide en dos partes:

1. La primera guía a las y los lectores en un proceso de **análisis exploratorio** en donde el objetivo es **definir** el problema en términos de las características de los datos de Twitter, **explorar** cómo se discute el tema de interés en redes so- ciales, **determinar** si el uso de datos de Twitter es o no pertinente dependiendo de la problemática que se desea examinar, y **evaluar** si la narrativa observada es de interés para llevar a cabo un estudio de mayor profundidad.

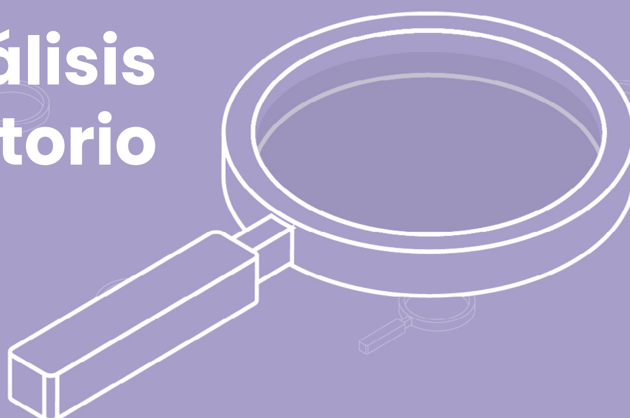


2. La segunda parte ofrece una descripción de cómo conducir un **análisis cualitativo a profundidad** enfatizando la reflexión sobre el proceso de interpretación, el cual inicia con la construcción del conjunto de datos a analizar, incluyendo la selección de la plataforma, las palabras clave que guían la búsqueda y recolección de tuits. Como resultado de este proceso las y los lectores podrán recopilar, interpretar, y sintetizar datos de Twitter para generar hallazgos que informen su investigación.

Proceso de análisis



Análisis exploratorio



Definición del problema

El primer paso del análisis exploratorio consiste en definir el problema que se desea explorar en Twitter. La definición consiste en identificar y enmarcar el problema que se explorará tanto para comunicar el alcance como el enfoque del análisis de datos de Twitter.

Herramientas:

Para acompañarte en el proceso de definición del problema, este kit incluye tres herramientas que sirven como guía. Se recomienda utilizarlas en el siguiente orden:

Plantilla de definición del problema:

Primero se recomienda llenar esta plantilla, la cual te permitirá recopilar información sobre el contexto y las comunidades involucradas en el problema, así como documentar la evidencia y los datos existentes. Al completar esta plantilla, reflexiona sobre los objetivos y las expectativas sobre utilizar datos de Twitter.

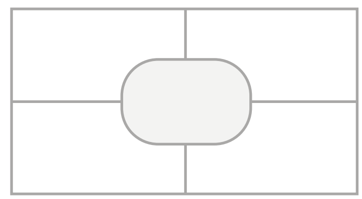
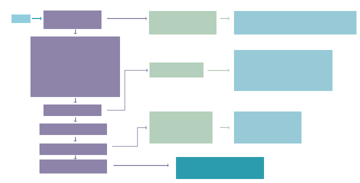


Diagrama de árbol de decisión:

Ahora que ya has recopilado más información sobre el contexto del problema que se pretende informar con datos de Twitter, te recomendamos utilizar el árbol de decisiones para evaluar si Twitter es o no una fuente apropiada.



Plantilla para documentación de palabras clave:

Utiliza esta plantilla para recopilar palabras clave que describan las características del problema, las poblaciones, comunidades u organizaciones involucradas. Así mismo, se recomienda la recolección de hashtags y cuentas de Twitter de aquellos actores que sean representativos del tópico a explorar. Si desconoces nombres de cuentas de Twitter, reflexiona sobre los factores que dan forma al problema que deseas examinar, piensa en palabras características de esos factores y con ellas busca cuentas de Twitter.

Exploración inicial de Twitter

El segundo paso del análisis exploratorio consiste en hacer una exploración inicial del problema en Twitter. Utilizando las palabras clave, cuentas de Twitter y hashtags recopilados en el paso anterior se realiza una búsqueda inicial con el objetivo de recopilar evidencia de cómo luce en Twitter el problema que se está examinando y así:

Determinar si el tópico es o no discutido en Twitter

Entender cómo se habla del tema

Identificar quienes son las comunidades, organizaciones, o actores que son parte de la conversación

Este paso es iterativo, es decir debe repetirse sucesivamente y en cada iteración sugerimos recopilar más palabras, cuentas de Twitter y hashtags para informar, dirigir y refinar las siguientes exploraciones.

Durante la exploración se recomienda también documentar los hallazgos de la búsqueda de las diferentes palabras clave con el objetivo de identificar actores, acciones, y discursos que nos informen sobre la narrativa de un problema. Se sugiere que al realizar cada búsqueda se documenten los resultados considerando los siguientes aspectos:

Tipo de contenido:	Tipo de discurso:	Actores involucrados:	Disponibilidad de datos:
Documenta el contenido que se comparte en Twitter en relación con las búsquedas realizadas. Por ejemplo, documenta si la tendencia es compartir noticias, fotografías, videos, etc.	Registra los tipos de discurso y discusiones observadas en los resultados. Por ejemplo, registra si los resultados arrojaron tuits que reflejen quejas, acciones locales, etc.	Documenta quien o quienes están discutiendo el tema. Registra el tipo de cuentas que comparten información relevante sobre el tema examinado. Por ejemplo, observa si las cuentas que publican o comparten información sobre el tema a examinar son de ciudadanos, de organizaciones no gubernamentales, del gobierno, de instituciones, activistas, periodistas, etc. En paralelo al registro, reflexiona sobre los patrones de los actores involucrados, por ejemplo que actores prevalecen en la mayoría de los resultados y qué actores están ausentes.	Documenta la disponibilidad de la información encontrada sobre el tema objetivo. El objetivo no es cuantificar los tuits sobre un tema específico sino más bien hacer una breve inspección sobre la frecuencia de tuits que reflejan el tema examinado.

En paralelo, se recomienda comparar los hallazgos preliminares obtenidos en Twitter con los datos de fuentes tradicionales y reflexiona si existe alguna información que esté disponible en las redes sociales que no se refleje en las típicas fuentes de datos.

Por último, durante la exploración, procura identificar los patrones y perspectivas que prevalecen en los resultados así como las que están ausentes.

Análisis | Exploración

Para acompañarte en el proceso de exploración este kit incluye herramientas computacionales y cualitativas que te guiarán y permitirán documentar los hallazgos iniciales. Se recomienda utilizar las herramientas en el siguiente orden:

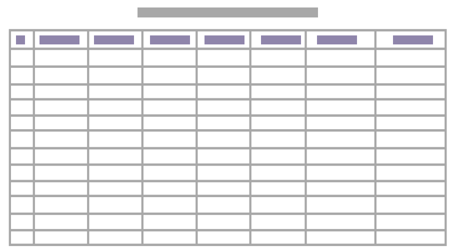
Plantilla para documentación de búsquedas exploratorias:

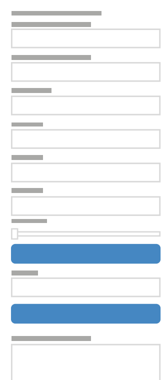
Te sugerimos utilizar esta plantilla para facilitar la documentación de los hallazgos de las búsquedas iniciales. El objetivo de esta plantilla es que se pueda llevar un control sobre cuáles combinaciones de palabras arrojan un contenido más interesante o apropiado que abone al entendimiento de problema que se está examinando.

Modulo 1:

Esta herramienta permite al usuario hacer búsquedas y descargas de tuits. Detalles sobre las capacidades de la herramienta e instrucciones sobre cómo utilizarlas se encuentran en la sección Herramientas de este manual.

Paso 2 | Exploración inicial de Twitter





Categorización de hallazgos iniciales

El tercer paso del análisis exploratorio consiste en hacer una categorización de los hallazgos iniciales. El objetivo es asociar el contenido de los tuits identificados en el paso anterior con la definición del problema que se está examinando. Para ello, es necesario revisar los hallazgos del paso dos y definir qué tipo de tuits son emblemáticos o codifican las dimensiones del tema investigado. Hacer esta asociación te permitirá decidir si Twitter es o no una fuente de información útil para tus objetivos y si es o no conveniente hacer una siguiente exploración analizando a mayor profundidad los datos de Twitter.

Durante la categorización se recomienda prestar atención a los siguientes aspectos:

Redundancia y patrones de información:

En este contexto, se entiende por redundancia la información que se repite en los datos de Twitter. En algunos casos, observar redundancia en los datos denota patrones de información, lo que a su vez puede contribuir a confirmar acciones de un grupo en específico o bien proporcionar nuevas perspectivas y conocimientos del tema que se está estudiando. La redundancia en los datos de Twitter puede tomar distintas formas, por ejemplo:

• El tipo de organizaciones o comunidades que más tuitean sobre un tema en específico

• El tipo de contenido que comparten las organizaciones

• La población o comunidad que más tuitea sobre un problema o situación, por ejemplo: adolescentes, activistas, trabajadores del gobierno, etc.

• El tipo de peticiones que un grupo en específico tuitea
Buscar estos patrones durante las búsquedas exploratorias permitirá definir mejor los temas y perspectivas que se desean ahondar más y aquellos que se deseen omitir.

Información no documentada en fuentes de datos tradicionales:

Como se ha mencionado anteriormente, una de las fortalezas de los datos de Twitter es que reflejan perspectivas que no siempre están documentadas en fuentes de información tradicionales. Por lo tanto se recomienda evaluar si los datos encontrados en Twitter evidencian actores, poblaciones o estrategias de las comunidades afectadas que anteriormente no se conocían o habían considerado.

Análisis | Exploración

Paso 3 | Categorización de hallazgos iniciales

Este kit incluye dos plantillas de trabajo que te acompañarán en el proceso de categorización de hallazgos iniciales. Se recomienda utilizarlas en el siguiente orden:

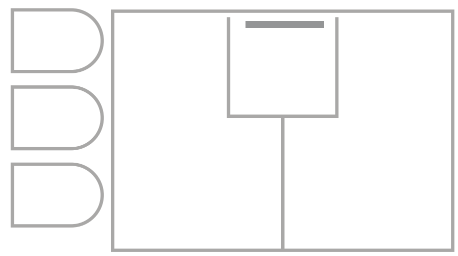
Plantilla de documentación de hallazgos y observaciones:

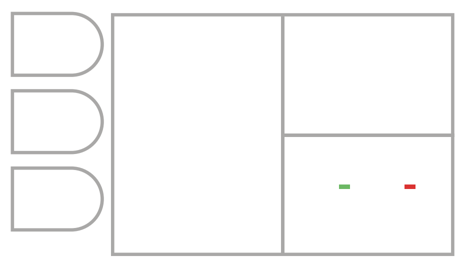
Para completar esta plantilla, utiliza como insumo el contenido de la plantilla para documentación de búsquedas exploratorias del paso dos. Para hacer la categorización de tuits puedes usar un diagrama de afinidad o bien puedes simplemente identificar las categorías que se repitan en la plantilla y capturar los hallazgos que te resulten más relevantes.

Se recomienda registrar ejemplos de tuits que sean ilustrativos de los hallazgos, ya que esto facilita la comunicación de resultados.

Plantilla de documentación de siguientes pasos:

Utiliza esta plantilla para reflexionar sobre las observaciones recopiladas en las búsquedas exploratorias y evalúa si la narrativa observada en Twitter es de interés para realizar un análisis más profundo. Al finalizar de completar esta plantilla deberás decidir si es o no recomendable continuar usando datos de Twitter para informar el problema examinado.





Análisis cualitativo

Recolección de datos

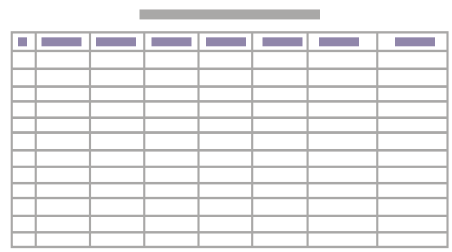
El primer paso del análisis cualitativo a profundidad consiste en hacer una recopilación de datos más extensa que la del análisis exploratorio. Para la recopilación de datos, se sugiere utilizar la herramienta del módulo 1. La selección de datos recopilados deberá guiarse por los hallazgos y aprendizajes obtenidos durante la búsqueda exploratoria.

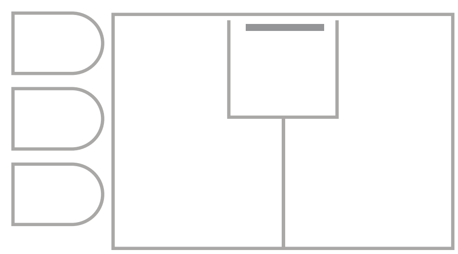
Herramientas:

Se sugiere guiar la recolección de datos empleando las hojas de trabajo previamente descritas como son:

1. Plantilla para documentación de búsquedas exploratorias

2. Plantilla de documentación de hallazgos y observaciones





Selección de conjunto de datos

El segundo paso del **análisis cualitativo a profundidad** consiste en seleccionar un conjunto de datos de Twitter más específico para conducir un análisis a profundidad.

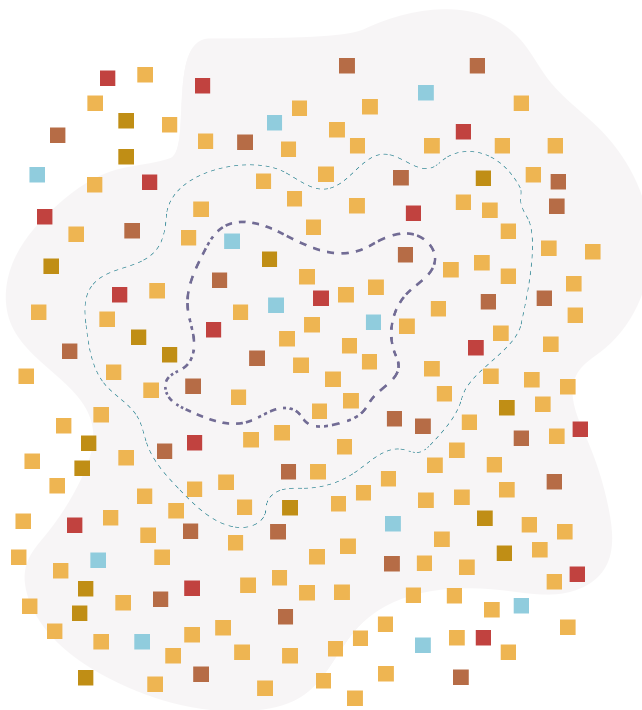
Una vez recolectados los tuits con la herramienta del módulo 1, utiliza la herramienta del módulo 2 para identificar los tuits que contienen contenido similar al problema que se está examinando.

Revisa la sección de Herramientas computacionales de este manual para una descripción detallada de cómo utilizar la herramienta del módulo 2.

Después de usar la herramienta del módulo 2, es posible que el número resultante de tuits sea aún bastante alto, por ejemplo más de 1000 tuits. Si este fuera el caso, se recomienda reducir todavía más el número de tuits para el análisis cualitativo con una segunda iteración, seleccionando aleatoriamente tuits lo cual puede hacerse uti- lizando la función aleatoria de Excel. El número de tuits seleccionados aleatoriamente dependerá de la capacidad que se tenga para analizarlos manualmente. Si se cuenta con un tiempo reducido, se sugiere seleccionar entre 20 y 50 tuits por tema para ana- lizarlos a profundidad.

Alternativamente, se puede hacer el segundo filtrado de tuits seleccionando únicamente aquellos que incluyan alguna palabra o término que haga referencia a zonas geográficas (estados o alcaldías), actores, palabras o hashtags específicos.

La selección de conjunto de datos es un paso iterativo y el objetivo es reducir los tuits para hacer un análisis cualitativo a profundidad. Por lo tanto, el resultado de este paso debe ser un número manejable de tuits que puedan analizarse a detalle.



Análisis | Cualitativo

Lectura y documentación de información adicional

El tercer paso del análisis cualitativo a profundidad consiste en leer cada uno de los tuits seleccionados en el paso anterior y en paralelo documentar información adicional de los tuits con el fin de comprender con mayor exactitud el contexto en el que se ha generado este contenido.

Para documentar cualquier información adicional de los tuits se recomienda emplear las plantillas previamente utilizadas como son:

1. Plantilla para documentación de búsquedas exploratorias

2. Plantilla de documentación de hallazgos y observaciones

Como ya se ha mencionado, el objetivo de estas plantillas es registrar información suplementaria de los tuits, esto con el fin de guiar su análisis e interpretación. La información adicional dependerá del tipo de problema que se está examinando. Sin embargo, se recomienda por lo menos cubrir los siguientes aspectos:

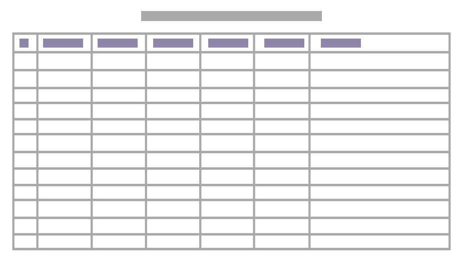
Actores involucrados:

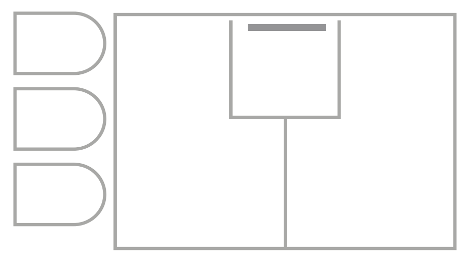
Persona u organización que publica el tuit y las reacciones que recibe. En estudios anteriores se ha observado que para interpretar el contenido de los tuits es necesario considerar a la persona que lo publica y su relación con las contrapartes involucradas. Así mismo el tipo de reacciones que reciben los tuits ayudan a valorar mejor el im- pacto que estos tienen. Por ejemplo, el número de Me gusta o el número de retweets.

Documentación de patrones:

Si bien las categorías dependen del tipo de problema que se está examinando recomendamos documentar patrones de problemáticas, necesidades, respuestas e iniciativas colectivas reportadas en Twitter. Conforme se avance tanto en la categorización como en la documentación de tuits, es necesario prestar particular atención a los puntos en común, distinciones y relaciones entre actores que el propio contenido de Twitter revela.

Paso 3 | Lectura y documentación de información adicional





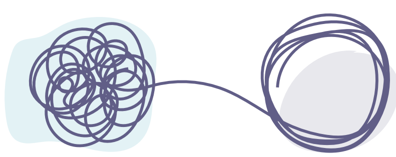
Categorización de tuits

El cuarto paso del análisis cualitativo a profundidad consiste en identificar las diversas expresiones o formas en que se manifiesta en Twitter el problema que se está investigando. Para ello, es necesario categorizar el conjunto de tuits que se definió y documentó en los pasos anteriores.

La categorización debe hacerse siguiendo una clasificación previamente definida con contrapartes involucradas en el proyecto o bien de forma colaborativa, esto con el fin de hacer una interpretación de los datos de forma consensuada. Para definir las categorías se recomienda leer y extraer las categorías que representen el tópico examinado.

Una vez definidas las categorías se recomienda organizar los tuits con el fin de identificar patrones, para ello se sugiere hacer un diagrama de afinidad, el cual esta descrito en la sección de herramientas cualitativas de este manual.

Por último, cabe señalar que la categorización de tuits es un paso iterativo y debe repetirse hasta que se hayan definido suficientes categorías que describan satisfactoriamente el problema que se esta examinando. En la imagen de la derecha se muestran dos ejemplos de tuits que se clasificaron como evidencia del capital social de puente [1] durante la exploración del rol de los lazos sociales en la reacción de comunidades ante la pandemia de COVID-19 [2].



[1] El capital social de puente es aquel que conecta a diferentes tipos de personas y grupos (étnicos, sociales, de género, política o regional, etc.)

[2] Munguia, J., & Alvarado, A. (2021, December 2). Exploring the role of social ties in a community's reaction to the COVID-19 pandemic through Twitter: El PNUD en México. UNDP

Tweet de un ciudadano que muestra a un grupo de vecinos donando alimentos

Tweet que describe la iniciativa de un restaurante que donó 150 cenas para el personal médico

Síntesis

El quinto y último paso del análisis cualitativo a profundidad consiste en sintetizar la información recopilada y categorizada en los pasos anteriores. El objetivo de este paso es darles sentido a los tuits recopilados buscando patrones y correlaciones en los datos para responder a las preguntas iniciales que motivaron el estudio. El resultado de este paso debiera ser una caracteriza- ción del problema que se está examinando.

Como ya se ha mencionado anteriormente, los datos de las plataformas de redes sociales siempre están incompletos, debido a sus limitaciones en términos de representación. Por lo tanto, las observaciones recopiladas a través de Twitter nunca serán exhaustivas de las problemáticas sociales que se están examinando. Sin embargo, los datos provenientes de Twitter pueden ofrecer evidencia de reacciones, respuestas y perspectivas colectivas de problemas sociales que de otro modo resultaría im- posible de identificar. Teniendo en cuenta tales limitaciones y restricciones sugerimos que lo más apropiado para dar sentido a los datos recopilados en Twitter es seguir un enfoque holístico en la interpretación de los datos y entender la problemática que se está examinando. Este enfoque holístico puede tomar diversas formas, pero el objetivo central es caracterizar los distintos componentes como son actores, geografías, perspectivas y temáticas que delinean el problema examinado. Por lo tanto, esta caracterización va estrictamente ligada con cómo se ha enmarcado inicialmente el problema y las preguntas que se buscan responder.

Utilizando como insumos los resultados de los pasos 3 y 4 identifica las posibles conexiones entre las categorías. Después resume los hallazgos con un enfoque holístico, como se describe en los párrafos anteriores. A continuación se enlistan algunas sugeren- cias de las posibles formas de organización de los datos:

Áreas geográficas: Haciendo observaciones y comparaciones que tomen como eje las geografías de los actores involucrados. Es decir, si los datos así lo permiten, centrar el análisis en las particularidades, diferencias, capacidades y necesidades de comu- nidades, dependiendo de su localidad geográfica, esto puede ser por alcaldía, ciudad o estado.

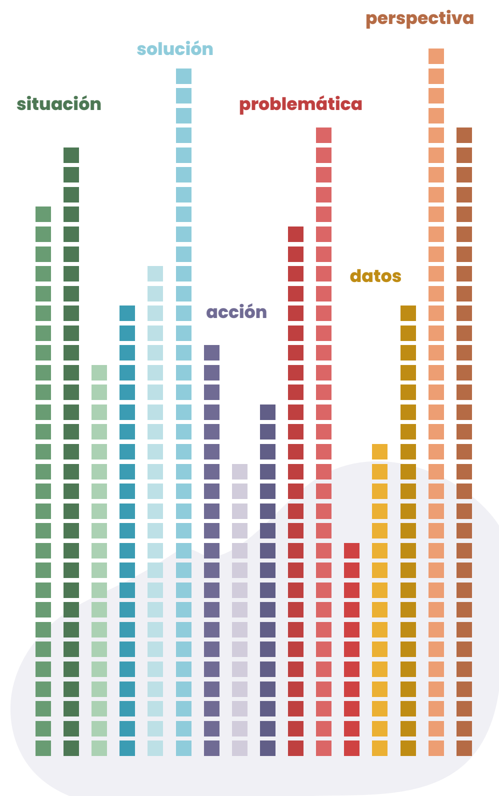
Tipo de solución: Soluciones que responden a necesidades locales pero que también dejan de manifiesto el aprovechamiento de las capacidades de las comunidades.

Acción - situación: Observaciones o hallazgos que ejemplifiquen una acción o respuesta colectiva ante una situación particular.

Problemáticas locales: Hallazgos que ilustren problemáticas arraigadas a las particularidades de las comunidades. Independientemente de la organización del análisis, lo que se recomienda es seguir un enfoque centrado en:

• Identificar perspectivas, capacidades y conocimientos dentro de la comunidad que no se habían considerado anteriormente

• Responder las preguntas iniciales que motivaron el estudio



Recomendaciones generales para el análisis de datos de Twitter

Decidir si es apropiado usar datos de Twitter u otras plataformas sociales depende de 1) el problema que los profesionales pretenden examinar, 2) cómo planean usar los datos y 3) las condiciones de las comunidades examinadas.

Recomendaciones para decidir que problemas son apropiados para examinar utilizando datos de Twitter

Si bien es posible buscar cualquier cantidad de palabras clave para averiguar sobre diferentes temas en Twitter, los problemas con las siguientes características son más adecuados para ser examinados utilizando datos de Twitter.

- Crisis locales en curso que afectan al grueso de la población

- Situaciones en las que los ciudadanos solicitan cambios al gobierno o a cualquier otra autoridad porque están insatisfechos con las condiciones sociales o económicas y las decisiones políticas.

- Desastres naturales, crisis provocadas por el hombre, como guerras y ataques terroristas, y crisis de enfermedades: las plataformas de redes sociales se han convertido en una parte rutinaria de la respuesta a la crisis y una infraestructura formalizada para diversas tareas, como recopilar relatos de testigos, monitorear los medios y proporcionar actualizaciones durante las respuestas a la crisis.

- Crisis relacionadas con violaciones de derechos humanos.

Los problemas que no son aptos para ser examinados con datos de Twitter son aquellos problemas sociales considerados sistémicos como la pobreza, la corrupción o el desempleo. Se recomienda evitar el análisis de este tipo de temas porque el discurso y los datos en Twitter reflejan principalmente tendencias y situaciones en curso en lugar de prácticas normalizadas en un contexto específico. Además de que este tipo de problemas suelen ser en extremo complejos y son el resultado de múltiples factores, lo que complica aún más su estudio utilizando datos de Twitter.

La segunda categoría de problemas que no es aconsejable examinar con los datos de Twitter son aquellas situaciones relacionadas con un grupo de población que corresponde a un género, grupo de edad o profesión específico. Debido a la naturaleza de la plataforma, es difícil determinar esas características de las y los usuarios a partir de los datos que comparten en Twitter. Por ejemplo, si el interés es conocer la perspectiva de mujeres adolescentes (de 13 a 19 años) o personas con una profesión específica sobre algún tema, entonces Twitter no será una fuente recomendada.

Recomendaciones generales para el análisis de datos de Twitter

Sugerencias para definir cómo utilizar los datos de Twitter

Un segundo aspecto para considerar al determinar qué temas son los más apropiados para examinar usando datos de Twitter se relaciona con cómo los profesionales de desarrollo planean usar los hallazgos de sus exploraciones en Twitter, para ello se recomienda reflexionar sobre el tipo de argumentos que pretenden presentar y la evidencia que podrían necesitar para respaldarlos. En este sentido, es imperativo distinguir entre usar la evidencia de Twitter para informar a otras partes interesadas (por ejemplo, la policía y otras organizaciones sin fines de lucro) o informar al equipo dentro de la organización para definir los próximos pasos.

Independientemente de la definición de evidencia, Twitter puede ser una fuente adecuada de información en caso de que los profesionales pretendan comprender y recopilar datos sobre los aspectos de cualquier comunidad que se incluyan en las siguientes categorías:

Experiencias y estrategias de las personas durante las crisis: Comprender con mayor detalle cómo las comunidades experimentan, reaccionan y resuelven los problemas locales.

Actores involucrados: Identificar actores y actores que contribuyan a la resolución de crisis locales que no fueron identificados previamente.

Discurso: Aprender sobre las múltiples opiniones y experiencias de una crisis en curso o pasada, e identificar las perspectivas de las comunidades que a menudo no se consideran.

Por el contrario, no se recomienda utilizar los datos de Twitter si el fin último es alguno de los siguientes:

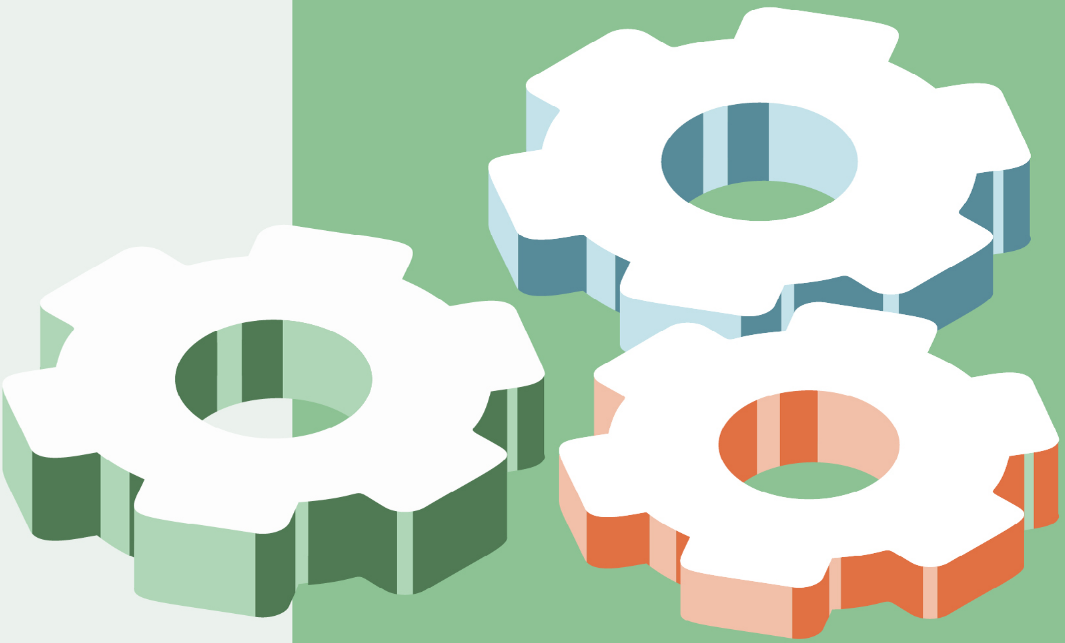
1. Cuantificar las condiciones sociales, culturales, económicas o políticas de una población: Twitter no debe utilizarse como única fuente de información para cuantificar fenómenos sociales. Como se explicó anteriormente, debido al formato de los datos de Twitter, siempre están incompletos y nunca son exhaustivos de las experiencias de las personas.

2. Responder inquietudes exactas de una población determinada: como ya se ha mencionado, no se recomienda utilizar únicamente datos de Twitter para recopilar datos de comunidades clasificadas por edad, género, profesión o ubicación.

Consideraciones sobre las Condiciones de las Comunidades

Un último conjunto de consideraciones que se deben hacer al decidir los temas a analizar usando Twitter está relacionado con las condiciones de las comunidades examinadas con respecto a su calidad de acceso a Internet, acceso a teléfonos inteligentes y sus preferencias de plataformas de redes sociales. Considerar estos tres aspectos es fundamental ya que determinan quién accede a Twitter, su frecuencia e incluso sus habilidades para organizarse utilizando tecnologías digitales. Si bien cumplir con estas condiciones es muy deseable, si no se cumplen es posible examinar los datos de Twitter siempre y cuando se tomen las medidas necesarias para contrarrestar los posibles sesgos que el limitado acceso a internet y teléfonos inteligentes pueda imponer a las comunidades estudiadas.

04 HERRAMIENTAS



HERRAMIENTAS

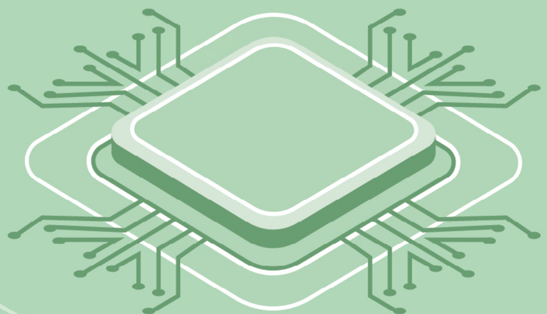
HERRAMIENTAS COMPUTACIONALES

Modulo 1:
Herramienta para búsqueda exploratoria y recolección de datos

Modulo 2:
Herramienta para el análisis de datos

HERRAMIENTAS CUALITATIVAS

- Diagrama de afinidad
- Plantilla de definición del problema
- Plantilla para documentación de palabras clave
- Plantilla para documentación de búsquedas exploratorias
- Plantilla de documentación de hallazgos y observaciones
- Plantilla de documentación de siguientes pasos



Herramientas Computacionales

HERRAMIENTAS COMPUTACIONALES

MÓDULO 1

Herramienta para buscar y descargar tuits

En este módulo se incluye una herramienta que le permitirá al usuario hacer una búsqueda exploratoria en Twitter para examinar el tipo de contenido asociado con el tema que le interese examinar, así como descargar un determinado de número de tuits.

MÓDULO 2

Herramienta para el análisis de datos

Este módulo contiene una herramienta de procesamiento de lenguaje natural (NPL por sus siglas en inglés) que facilita el análisis de una enorme cantidad de tuits de forma inmediata, reduciendo así la sobrecarga de examinar tuits manualmente.

El propósito de la herramienta es filtrar tuits e identificar aquellos con contenido similar al problema que se está examinando.

Antes de usar este módulo te recomendamos visitar el módulo 1 para generar correctamente los insumos requeridos para utilizar esta herramienta.

Módulo 1

Descripción de la herramienta

El objetivo de esta herramienta es facilitar la exploración inicial de datos en Twitter y su recolección.

La interfaz permite realizar búsquedas de datos de tuits con los siguientes criterios:

1. Contenido del tuit: En esta sección, puedes definir palabras clave y hashtags para filtrar la búsqueda de contenido.

a. Evita utilizar frases y el uso de comillas (“ ”)

b. Es posible realizar búsquedas con múltiples palabras, para ello solo es necesario separarlas por comas.

c. También es posible definir palabras que deseas omitir en la búsqueda de datos. Te recomendamos utilizar este filtro para obtener una búsqueda más precisa

2. Cuenta de Twitter: Esta opción te permite filtrar los tuits por usuarios específicos de Twitter

3. Localización: Este control te permite definir la ubicación específica de los tuits que deseas examinar. Es posible filtrar la búsqueda por país, estado, o incluso alcaldía.

a. Este control te permite definir la ubicación específica de los tuits que deseas recolectar.

b. Recuerda que es posible filtrar la búsqueda por país, estado, o incluso alcaldía.

4. Fechas: Este control te permite definir el periodo de tiempo del cual deseas recolectar datos.

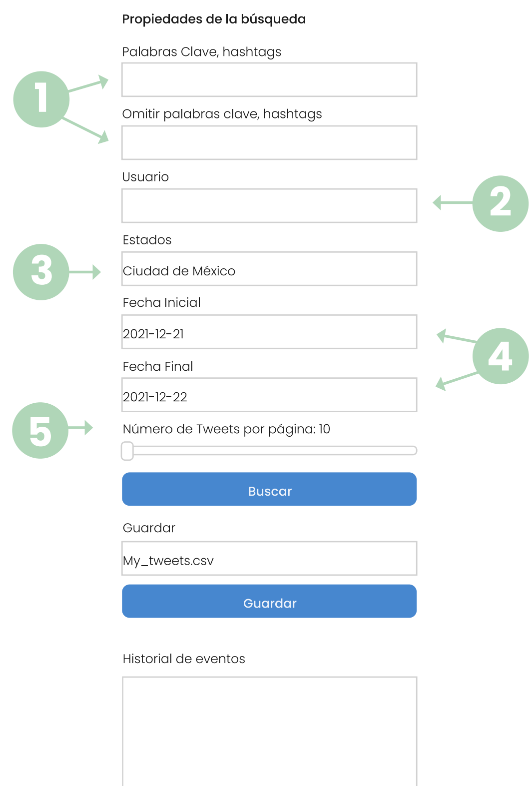
a. Utiliza la opción de calendario para definir el periodo de tiempo del cual deseas buscar datos. Es posible buscar y recopilar datos desde marzo del 2006.

b. Te recomendamos buscar y guardar datos correspondientes a distintos periodos de tiempo para comparar e identificar cambios en el discurso sobre un mismo tema.

5. Número de tuits: Esta opción te permite fijar el número máximo de tuits que se mostrarán en pantalla y que podrás descargar.

a. Es posible desplegar de 1 a 500 tuits por búsqueda dependiendo del tipo de usuario

6. Idioma: Por default el idioma de los tuits está predefinido al español



Módulo 1

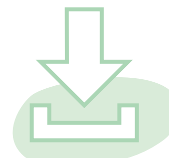
Descripción de la herramienta

Una vez realizada la búsqueda con la herramienta del módulo 1, los resultados se desplegarán en la pantalla y al hacer clic en el botón de guardar, los datos se salvarán en un archivo .CSV con el nombre deseado.

La interfaz permite además seleccionar los tuits que desean obtenerse, en caso de no necesitar todos los tuits resultantes de la búsqueda. Los datos de Twitter que se descarguen pueden ser analizados con la herramienta del módulo dos.

Por último, al descargar los tuits se incluye la siguiente información:

Como recordatorio, el archivo de tuits generado por la herramienta del módulo 1 incluye los siguientes campos:



1. **username:** Nombre de usuario de la cuenta de Twitter

2. **text:** Contenido del tuit

3. **created_at:** Fecha y hora de creación del tuit.
   El formato de la fecha es: año - mes - día

4. **link:** URL del tuit

5. **place_full_name:** Si está disponible nombre del estado, ciudad, o alcaldía.

6. **place_name:** Nombre del lugar

7. **place_type:** Especifica el tipo particular del lugar como el nombre de una ciudad o un punto de interés.

8. **country:** Nombre del país

	A	B	C	D	E	F	G	H
1	username	text	created_at	link	place_full_name	place_name	place_type	contry
2	ellaesadriana	Dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat	2022-01-10 21:00:53+00:00	https://twitter.com/twipsum/status/13007058604	Azcapotzalco, Distrito Federal	Azcapotzalco	city	México

Encabezados del archivo de tuits

Módulo 2

Descripción de la herramienta

El propósito de esta herramienta es filtrar tuits e identificar aquellos con contenido similar al problema que se está examinando. Para utilizar esta herramienta se requiere de dos archivos:

Archivo 1: Un archivo que contenga tuits

Archivo 2: Un conjunto de enunciados que describen el problema que se está investigando

En términos prácticos, la herramienta evalúa la similitud de los tuits con los enunciados con los que se alimenta, dando como resultado aquellos tuits que son semánticamente similares a los enunciados dados.

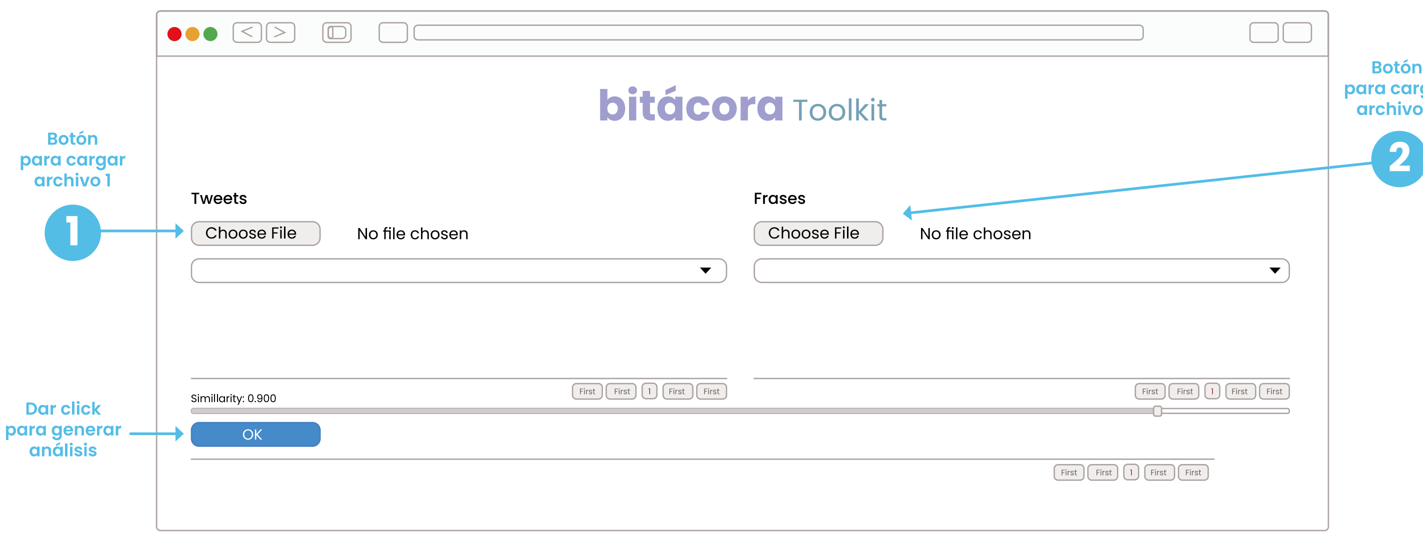


Imagen de la interfaz

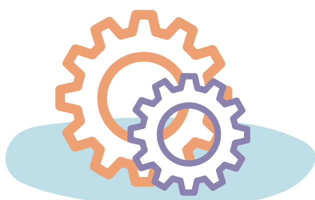
Módulo 2

Descripción de la herramienta

Este módulo contiene una herramienta de procesamiento de lenguaje natural (NPL por sus siglas en inglés) que facilita el análisis de una enorme cantidad de tuits de forma inmediata, reduciendo así la sobrecarga de examinar tuits manualmente.

Esta herramienta utiliza el algoritmo word2vec, el cual consiste primero en codificar texto como vector para después examinar el significado semántico del texto y hacer comparaciones vectoriales para identificar similitudes. La herramienta utiliza el modelo Spanish Billion Word Corpus and Embeddings (SBWCE) que consta de casi 1.500 millones de palabras semánticamente relacionadas en español, lo que a su vez nos informa del contexto de las palabras.

En esta sección del manual se describe el proceso para utilizar la herramienta para el análisis de datos.



Pasos

1. Revisión del archivo de tuits previamente recolectados (archivo 1)
2. Preparación de lista de frases que caracterizan el problema a examinar (archivo 2)
3. Análisis utilizando la herramienta

Módulo 2

Revisión del archivo de tuits

Para utilizar la herramienta del módulo 2, además de la lista de frases, es necesario tener un conjunto de datos de tweets del que la herramienta extraerá aquellos que sean semánticamente similares a las frases.

Por lo tanto, una vez generada la lista de frases, el siguiente paso es asegurarse de generar un archivo de tuits utilizando el módulo anterior. Idealmente, los tuits deben corresponder al período de tiempo que se está examinando, y el archivo debe generarse con base a las palabras clave, cuentas de Twitter, y ubicación geográfica que reflejen el tema a examinarse.

Como recordatorio, el archivo de tuits generado por la herramienta del módulo 1 incluye los siguientes campos:

1. username: Nombre de usuario de la cuenta de Twitter

2. text: Contenido del tuit

3. created_at: Fecha y hora de creación del tuit.
El formato de la fecha es: año - mes - día

4. link: URL del tuit

5. place_full_name: Si está disponible nombre del estado, ciudad, o alcaldía.

6. place_name: Nombre del lugar

7. place_type: Especifica el tipo particular del lugar como el nombre de una ciudad o un punto de interés.

8. country: Nombre del país

	A	B	C	D	E	F	G	H
1	username	text	created_at	link	place_full_name	place_name	place_type	contry
2	ellaesadriana	Dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat	2022-01-10 21:00:53+00:00	https://twitter.com/twipsum/status/13007058604	Azcapotzalco, Distrito Federal	Azcapotzalco	city	México

Encabezados del archivo de tuits

Módulo 2

Preparación de lista de frases

¿Qué es la lista de frases?

El primer paso consiste en armar una lista de frases que caractericen el problema. Estas frases pueden ser extractos de tuits que ilustren múltiples discursos desde distintos ángulos y perspectivas de los actores y organizaciones involucradas del tema a examinar. Alternativamente, las frases pueden ser también extractos de entrevistas, diálogos, respuestas a encuestas, o de cualquier otra fuente. Independientemente de su origen, lo importante es recolectar contenido que refleje el tema que se desea investigar en Twitter, y en la medida de lo posible desde distintas perspectivas.

¿Como hacer la lista de frases?

A continuación se enlistan una serie de recomendaciones para generar las frases que la herramienta del módulo 2 utilizara:

• Generar oraciones completas (sujeto, verbo, predicado) y cortas

• Elimina preposiciones, y detalles o referencias a otros contextos o tópicos • Evita oraciones largas

• Las frases deben incluir contenido del tema a examinar

Si extraes frases de tuits te sugerimos:

• Extraer el texto del tuit y transformarlo en oraciones completas

• Cuando el tuit incluya una mención, por ejemplo, @Alc_Iztapalapa, sugerimos extraer el nombre propio, en este caso "Iztapalapa".

• Evita copiar y pegar los tuits, procura siempre limpiar el texto y asegurarte de que las frases estén completas y coherentes.

• Omite las oraciones o expresiones que no tengan ningún contenido relacionado con el tema a investigar.

Formato del archivo

• La lista de frases debe estar en una columna (como se muestra en la imagen) y el título de la columna debe ser "text".

• La lista debe guardarse en formato .CSV

	A
1	text
2	Acerca la venta de frutas y verduras a distintas colonias
3	Venta de frutas y verduras a precio solidario
4	Alimentos para quien perdio su empleo a causa del covid 19
5	Aportan a la sociedad con entregas a domicilio
6	Apoya a familiares
7	Apoyo a los agricultores por medio de la compra justa de sus productos
8	Ayuda a los adultos mayores sin costo
9	Ayuda a trabajadoras sexuales
10	Ayuda a usuarios de drogas
11	Ayudan a la gente que trabaja en la calle sin sueldo fijo
12	Ayudan a la sociedad para que los vecinos no salgan
13	Ayudan con entregas a domicilio para que los adultos mayores no salgan
14	Ayudan pagando un alimento extra
15	Ayudar a productores y comerciantes de hortalizas
16	Ayudar a personas violentadas en sus hogares
17	Brinda alimentos a comunidad LGBTTI
18	Brinda alimentos a trabajadores de salud de manera gratuita
19	Brinda tarjetas con saldo
20	Proporcionan vales

Ejemplo del formato del archivo de la lista de las frases

Módulo 2

Preparación de lista de frases

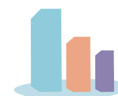
¿Como hacer la lista de frases?

Algunos ejemplos de cómo transformamos texto de distintas fuentes en frases listas para ser utilizadas con la herramienta del módulo 2:

Fuente	Texto Original	Texto Limpio
Twitter	Alrededor de unos 60 ejidatarios del pueblo de San Gregorio, en Xochimilco, se unieron para entregar sus productos al domicilio de quien lo solicite en la #CDMX. Aquí encontrarás el número para realizar pedidos:	Ejidatarios de San Gregorio, Xochimilco entregan productos a domicilio.
Twitter	De la chinampa es un grupo de agricultores de Xochimilco que comercializa sus productos vía online, actualmente siguen distribuyendo sus productos, checa las fechas de entrega.	Agricultores comercializan productos vía online
Encuesta	Esta iniciativa consiste en que se entregan 250 comidas gratuitas al día, cocinadas cada día por un chef diferente, a pacientes y sus familiares que reciben tratamiento en el Hospital Centro Médico Siglo XXI. Las comidas son también entregadas a personal médico del Instituto Nacional de Ciencias Médicas y de Nutrición y del Instituto Nacional de Enfermedades Respiratorias. El objetivo de la campaña es poder continuar con el apoyo a los agricultores por medio de la compra justa de sus productos y con la entrega de las comidas a pacientes, familiares y personal médico.	• Entrega de comidas gratuitas a pacientes y personal médico. • Apoyo a agricultores con la compra justa
Encuesta	Reparte despensas a la población vulnerable en la alcaldía de Xochimilco, así mismo a través de una alianza con la Plataforma México vs COVID inició el proyecto de Comedores Comunitarios donde reparten alrededor de 120 comidas diarias en la zona chinampera de la delegación.	• Reparte despensas a la población vulnerable • Comedores comunitarios reparten comidas

Módulo 2

Análisis utilizando la herramienta

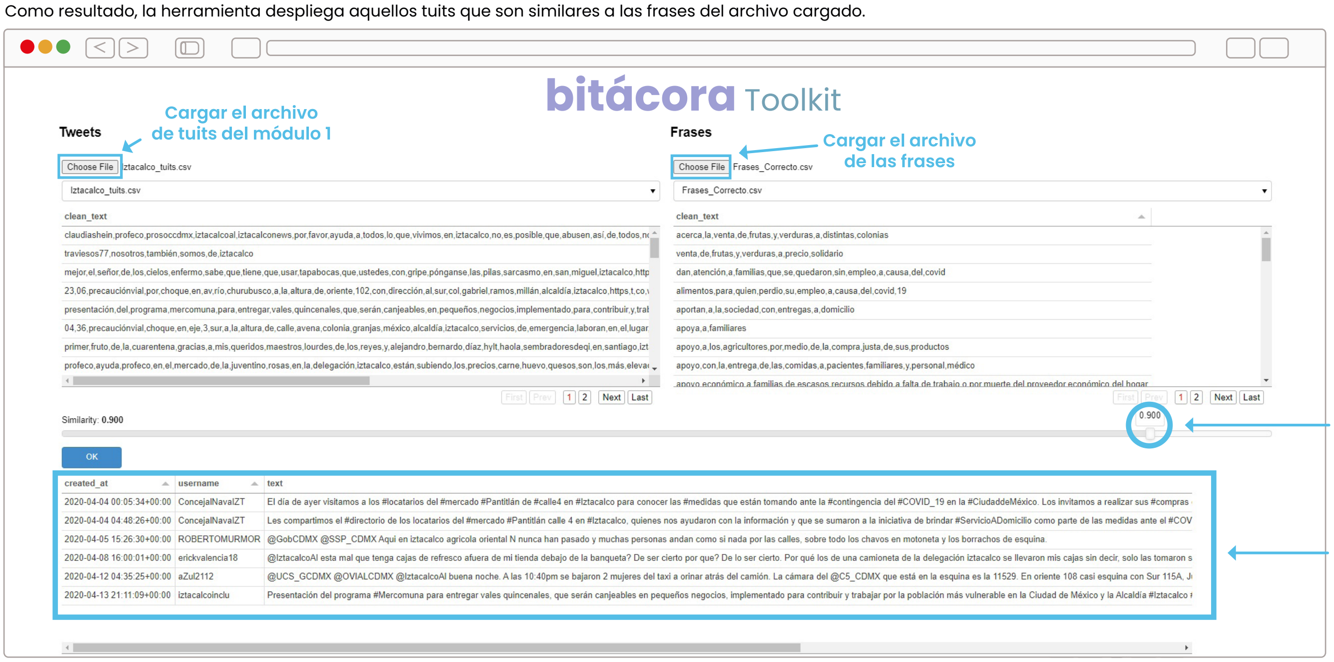


1. Una vez generados el archivo de la lista de frases y el archivo de tuits, ambos archivos se cargan en la herramienta del módulo 2.

2. Después con la barra de similarity, se determina el nivel de similitud que se busca entre los tuits y las frases. Los valores de similitud de la herramienta van de 0 (sin similitud) hasta 1 (mayor similitud).

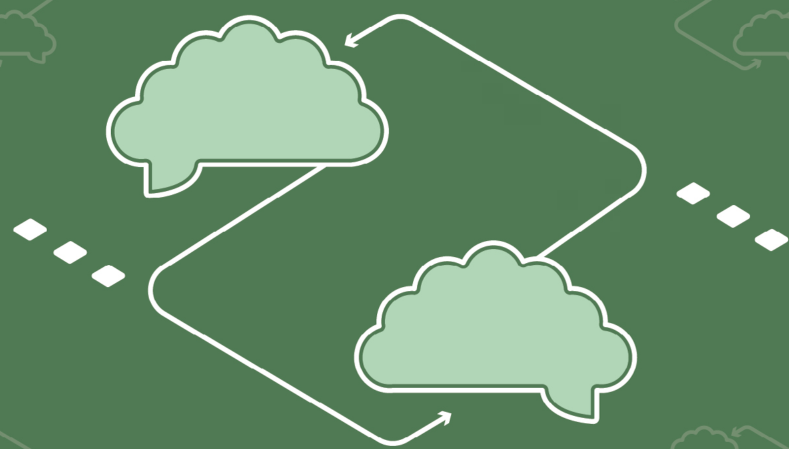
3. Por ultimo, se da click en OK y la herramienta inicia el análisis.

4. Como resultado, la herramienta despliega aquellos tuits que son similares a las frases del archivo cargado.



Interfaz del módulo 2

Herramientas cualitativas



Herramientas

Diagrama de afinidad

El diagrama de afinidad es un método que permite agrupar significativamente observaciones, ideas y datos, para identificar patrones y relaciones.

¿Cómo utilizar el diagrama?

En el contexto de este manual, el diagrama de afinidad puede ser utilizado para capturar, agrupar y sintetizar las observaciones recopiladas durante la revisión de tuits.

1. Registro de observaciones: Antes de utilizar el diagrama de afinidad, registra un promedio de 50 - 100 observaciones de los tuits. Cada observación debe estar en su propia nota y las notas pueden ser:

• Extractos de tweets

• Primeras impresiones sobre el contenido de los tuits, las temáticas de los tuits por ejemplo, necesidades, acciones, problemas, opiniones y perspectivas, la frecuencia con la que se discuten temas, el tipo de comentarios y conversaciones que se generan, etc.

• Observaciones sobre lo que les pareció interesante, lo que les sorprendió, lo que quisieran investigar un poco más

• Nuevos actores y organizaciones

2. Organización y categorización: Una vez generadas las notas, el equipo puede comenzar el proceso de organización y categorización buscando patrones y puntos en común. Las notas que comparten una intención, un problema o una cuestión similar se agrupan juntas. En el transcurso de este proceso, el equipo interpretara las notas, y conforme se avance en las rondas de análisis, surgirán patrones. Los patrones se perciben cuando los equipos clasifican los elementos en función de la similitud percibida, definiendo puntos en común que son inherentes pero no necesariamente obvios.

El resultado del diagrama de afinidad debiera ser una consolidación de las observaciones recolectadas que informen arquetipos agregados sintetizados. Idealmente, estos arquetipos o modelos proporcionaran una descripción detallada de las distintas comunidades, perspectivas o particularidades que integran la temática que se está examinando.

Finalmente, recomendamos utilizar los hallazgos y patrones identificados como punto de partida para informar siguientes pasos ya sea utilizando métodos adicionales como son entrevistas o encuestas para ahondar más en los temas que sean de interés.

Diagrama de afinidad

Nombre de la categoría:

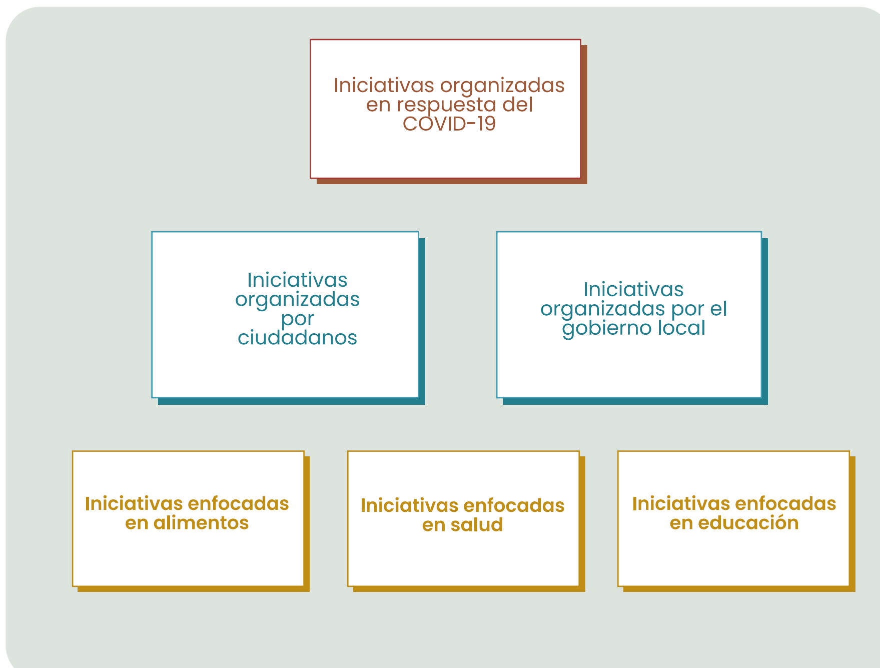
Palabra o frase corta utilizada para referirse a una categoría describiendo un aspecto relevante y general del tema que se está examinando.

Descripción de la categoría:

Breve texto que describa las principales características de la categoría reveladas por las notas amarillas. Pueden ser categorías de actores u organizaciones que llevan a cabo las acciones observadas y documentadas en las notas amarillas.

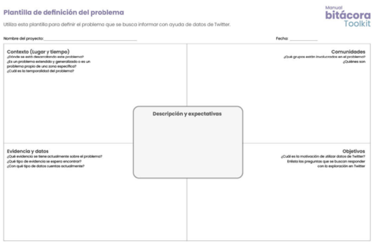
Evidencia asociada con la categoría:

Extractos de tuits u observaciones que ejemplifiquen las categorías. Estas observaciones son los componentes básicos del diagrama de afinidad.



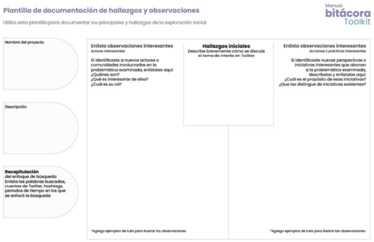
Ejemplo breve de una de las categorizaciones iniciales que hicimos durante el proyecto de inventario social

PLANTILLAS


Plantilla de definición del problema


Plantilla para documentación de palabras clave


Plantilla para documentación de búsquedas exploratorias


Plantilla de documentación de hallazgos y observaciones


Plantilla de documentación de siguientes pasos

Manual
bitácora
Toolkit

Plantilla de definición del problema

Utiliza esta plantilla para definir el problema que se busca informar con ayuda de datos de Twitter.

Manual **bitácora** Toolkit

Nombre del proyecto: ____ Fecha: ____________

Contexto (Lugar y tiempo)
¿Dónde se está desarrollando este problema?
¿Es un problema extendido y generalizado o es un problema propio de una zona específica?
¿Cuál es la temporalidad del problema?

Comunidades
¿Qué grupos están involucrados en el problema?
¿Quiénes son

Descripción y expectativas

Evidencia y datos
¿Qué evidencia se tiene actualmente sobre el problema?
¿Qué tipo de evidencia se espera encontrar?
¿Con qué tipo de datos cuentas actualmente?

Objetivos
¿Cuál es la motivación de utilizar datos de Twitter?
Enlista las preguntas que se buscan responder con la exploración en Twitter

Plantilla para documentación de palabras clave

Utiliza esta plantilla para registrar las palabras clave que caracterizan el problema que se desea explorar en Twitter.

Manual **bitácora** Toolkit

Nombre del proyecto: ______________________________ Fecha: __________

Descripción del problema	Organizaciones involucradas	Comunidades afectadas	Cuenta de Twitter de actores involucrados	Hashtags

Descripción:

Plantilla para documentación de búsquedas exploratorias

Utiliza esta plantilla para documentar tus observaciones de las búsquedas exploratorias.

Manual bitácora Toolkit

Nombre del proyecto: ______________________________

Fecha: __________

Documentación de búsquedas exploratorias

ID	Lugar	Temporalidad	Combinación de palabras clave	Número de resultados	Observaciones	Actores / Cuentas de Twitter	Ejemplos de tuits
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

Descripción:

Plantilla de documentación de hallazgos y observaciones

Utiliza esta plantilla para documentar los principales y hallazgos de la exploración inicial

Manual
bitácora
Toolkit

Nombre del proyecto

Descripción

Recapitulación
del enfoque de búsqueda
Enlista las palabras buscadas, cuentas de Twitter, hashtags, periodos de tiempo en los que se enfocó la búsqueda

Enlista observaciones interesantes
Actores interesantes

Si identificaste a nuevos actores o comunidades involucradas en la problemática examinada, enlístalas aquí:
¿Quiénes son?
¿Qué es interesante de ellxs?
¿Cuál es su rol?

Hallazgos iniciales
Describe brevemente cómo se discute el tema de interés en Twitter

Enlista observaciones interesantes
Acciones o prácticas interesantes

Si identificaste nuevas perspectivas o iniciativas interesantes que abonan a la problemática examinada, descríbelas y enlístalas aquí:
¿Cuál es el propósito de esas iniciativas?
¿Que las distingue de iniciativas existentes?

*Agrega ejemplos de tuits para ilustrar las observaciones

*Agrega ejemplos de tuits para ilustrar las observaciones

Plantilla para documentación de búsquedas exploratorias

Utiliza esta plantilla para documentar tus observaciones de las búsquedas exploratorias.

Manual
bitácora
Toolkit

Nombre del proyecto

Descripción

Recapitulación
del enfoque de búsqueda
Enlista las palabras buscadas, cuentas de Twitter, hashtags, periodos de tiempo en los que se enfocó la búsqueda

Relación entre hallazgos de Twitter y conocimientos previos

¿Cómo se relacionan o diferencian las observaciones encontradas en Twitter con lo que sabías del problema anteriormente?
¿Qué es nuevo? ¿Qué se repite?

*Agrega ejemplos de tuits para ilustrar las observaciones

Limitaciones

¿Qué limitaciones anticipas? Por ejemplo no encontrar información específica de un lugar o de una población en específico
¿Cuáles son algunas inquietudes o preocupaciones?

¿Sería interesante realizar una búsqueda más profunda?

Si

¿Cómo contribuyen los datos de Twitter a entender el problema a investigar?

No

¿Por qué?

*Agrega ejemplos de tuits para ilustrar las observaciones

REFERENCES

[1] A. Kube, S. Das, and P. J. Fowler, "Allocating interventions based on predicted outcomes: A case study on homelessness services," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 622–629, Jul. 2019.

[2] R. Richardson, J. Schultz, and K. Crawford, *Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice*, Mar. 2019.

[3] M. Willson, "Algorithms (and the) everyday," *Information, Communication & Society*, vol. 20, no. 1, pp. 137–150, 2017. eprint: https://doi.org/10.1080/1369118X.2016.1200645.

[4] O. Alvarado and A. Waern, "Towards algorithmic experience: Initial efforts for social media contexts," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–12, ISBN: 9781450356206.

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

[6] D. Citron and F. Pasquale, "The scored society: Due process for automated predictions," *Washington Law Review*, vol. 89, pp. 1–33, Mar. 2014.

[7] A. Chouldechova, D. Benavides-Prado, O. Fialko, and R. Vaithianathan, "A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, S. A. Friedler and C. Wilson, Eds., ser. Proceedings of Machine Learning Research, vol. 81, New York, NY, USA: PMLR, 2018, pp. 134–148.

[8] H. Ledford, "How facebook, twitter and other data troves are revolutionizing social science," *Nature*, vol. 582, no. 7812, pp. 328–330, 2020.

[9] "The powers and perils of using digital data to understand human behaviour," *Nature*, vol. 595, no. 7866, pp. 149–150, 2021.

[10] J. A. Pater, O. L. Haimson, N. Andalibi, and E. D. Mynatt, ""hunger hurts but starving works" characterizing the presentation of eating disorders online," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1185–1200.

[11] S. Chancellor, Z. Lin, E. L. Goodman, S. Zerwas, and M. De Choudhury, "Quantifying and predicting mental illness severity in online pro-eating disorder com-

munities," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1171–1184.

[12] S. Chancellor, Z. Lin, and M. De Choudhury, """ this post will just get taken down" characterizing removed pro-eating disorder social media content," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1157–1162.

[13] S. Chancellor, J. A. Pater, T. Clear, E. Gilbert, and M. De Choudhury, "# Thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 1201–1213.

[14] X. Wang, D. E. Brown, and M. S. Gerber, "Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information," in *2012 IEEE International Conference on Intelligence and Security Informatics*, IEEE, 2012, pp. 36–41.

[15] M. L. Williams *et al.*, "Policing cyber-neighbourhoods: Tension monitoring and social media networks," *Policing and society*, vol. 23, no. 4, pp. 461–481, 2013.

[16] danah boyd and K. Crawford, "Critical questions for big data," *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679, 2012. eprint: https://doi.org/10.1080/1369118X.2012.678878.

[17] A. Iliadis and F. Russo, "Critical data studies: An introduction," *Big Data & Society*, vol. 3, no. 2, p. 2 053 951 716 674 238, 2016. eprint: https://doi.org/10.1177/2053951716674238.

[18] C. DIgnazio and L. F. Klein, *Data Feminism*. MIT PRESS LTD, 2020.

[19] Y. A. Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society*. The MIT Press, 2019, ISBN: 0262039664.

[20] C. L. Borgman, *Big data, little data, no data: scholarship in the networked world*. The MIT Press, 2016.

[21] E. Aronova, C. v. Oertzen, and D. Sepkoski, "Introduction: Historicizing big data," *Osiris*, vol. 32, no. 1, pp. 1–17, 2017. eprint: https://doi.org/10.1086/693399.

[22] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," 2020. arXiv: 2012.05345 [cs.LG].

[23] M. Miceli, M. Schuessler, and T. Yang, "Between subjectivity and imposition: Power dynamics in data annotation for computer vision," vol. 4, no. CSCW2, Oct. 2020.

[24] M. Muller *et al.*, "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*, ser. CHI '19, Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–15, ISBN: 9781450359702.

[25] S. Passi and S. Jackson, "Data vision: Learning to see through algorithmic abstraction," ser. CSCW '17, Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 2436–2447, ISBN: 9781450343350.

[26] S. Passi and S. J. Jackson, "Trust in data science: Collaboration, translation, and accountability in corporate data science projects," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, Nov. 2018.

[27] K. H. Pine and M. Liboiron, "The politics of measurement and action," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15, Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 3147–3156, ISBN: 9781450331456.

[28] A. Alvarado Garcia and C. A. Le Dantec, "Quotidian report: Grassroots data practices to address public safety," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, Nov. 2018.

[29] A. Alvarado Garcia, M. J. Britton, D. M. Doshi, M. De Choudhury, and C. A. Le Dantec, "Data migrations: Exploring the use of social media data as evidence for human rights advocacy," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW3, Jan. 2021.

[30] S. Carlson and B. Anderson, "What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use," *Journal of Computer-Mediated Communication*, vol. 12, no. 2, pp. 635–651, Jan. 2007. eprint: https://academic.oup.com/jcmc/article-pdf/12/2/635/22317230/jjcmcom0635.pdf.

[31] O. S. Foundation, "Atrocidades innegables: Confrontando crimenes de lesa humanidad en méxico," *Open Society Foundation*, 2016.

[32] T. M. C. o. D. The Federation for Human Rights (FIDH), H. R. P. (CMDPDH), and T. C. C. of Human Rights of the Northeast (CCDH), "Mexico: Report on the alleged commission of crimes against humanity in baja california between 2006 and 2012," Oct. 2014.

[33] H. R. Watch, *World report 2019: Rights trends in mexico*, Jan. 2019.

[34] ——, "Neither rights nor security killings, torture, and disappearances in mexico's war on drugs," *Human Rights Watch*, Nov. 2011.

[35] K. Heinle, O. Rodríguez Ferreira, and D. A. Shirk, *Drug violence in mexico: Data and analysis through 2015 special report*, Apr. 2016.

[36] A. Alvarado Garcia, A. L. Young, and L. Dombrowski, "On making data actionable: How activists use imperfect data to foster social change for human rights violations in mexico," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, Dec. 2017.

[37] INEGI, *Encuesta nacional de victimización y percepción sobre seguridad pública (envipe).* 2016.

[38] ——, *Encuesta nacional de victimización y percepción sobre seguridad pública (envipe).* 2017.

[39] D. d. M. Vélez Salas and M. A. Vélez Salas, "Desapariciones forzadas e involuntarias: El registro estadístico de la desaparición: Delito o circunstancia?," Feb. 2017.

[40] INEGI, *Estadistica de defunciones generales. sintesis metodologica*, 2014.

[41] C. C. México, *Report enforced disappearance in mexico 2011*, Mar. 2011.

[42] G. C. Bowker and S. L. Star, *Sorting things out: classification and its consequences*. MIT Press, 2008.

[43] P. Mónaco Felipe, *Ya nadie publica eso: Derecho a la verdad, violencia contra la prensa y afectaciones a comunidades y colegas*, Aug. 2020.

[44] C. Kahn, *Number of journalists killed in mexico reaches 'historical high,' report says*, Dec. 2017.

[45] S. M. Ellis, *Mexico most deadly country for journalists in 2017*, 2017.

[46] *Iraq, syria deadliest countries for journalists*, Dec. 2017.

[47] *Rsf's 2020 round-up: 50 journalists killed, two-thirds in countries "at peace": Reporters without borders*, Dec. 2020.

[48] L. Castellanos, *México armado. 1943-1981*. ERA, 2007.

[49] C. E. G. Rangel Lozano and E. Sánchez Serrano, *MÉXICO EN LOS SETENTA: Guerra sucia o terrorismo de Estado? Hacia una política de la memoria.* Itaca / UAG, 2015.

[50] D. Wilkinson, *Mexico: Forced disappearance, an ongoing crime*, Nov. 2018.

[51] A. Alvarado Garcia and J. Munguia, *Exploring the role of social ties in a community's reaction to the covid-19 pandemic through twitter: El pnud en méxico*, Nov. 2021.

[52] G. Rios Landa, A. Cervantes, and D. Torres, *Identificando espacios públicos (más) seguros para las mujeres en ciudad de méxico. parte 2: Del análisis cuantitativo al cualitativo: El pnud en méxico*, Feb. 2022.

[53] U. Mexico, *Lanza pnud inventario social covid-19 para identificar iniciativas que fortalezcan respuesta cívica de la cdmx: El pnud en méxico*, May 2020.

[54] Z. Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. YALE University Press, 2017.

[55] J. Aronson, *Challenging impunity with user generated content*, Mar. 2018.

[56] J. D. Aronson, "Crowd-training machine learning systems for human rights abuse documentation," in *AAAI 2014*, 2014.

[57] H. I. Evans *et al.*, "Facebook in venezuela: Understanding solidarity economies in low-trust environments," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–12, ISBN: 9781450356206.

[58] D. Dailey and K. Starbird, "Social media seamsters: Stitching platforms and audiences into local crisis infrastructure," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17, Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 1277–1289, ISBN: 9781450343350.

[59] M. Wong-Villacres, C. M. Velasquez, and N. Kumar, "Social media for earthquake response: Unpacking its limitations with care," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, Dec. 2017.

[60] N. Andalibi, P. Ozturk, and A. Forte, "Sensitive self-disclosures, responses, and social support on instagram: The case of depression," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17, Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 1485–1500, ISBN: 9781450343350.

[61] N. Andalibi, O. L. Haimson, M. D. Choudhury, and A. Forte, “Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media,” *ACM Trans. Comput.-Hum. Interact.*, vol. 25, no. 5, Oct. 2018.

[62] C. Crivellaro, R. Comber, J. Bowers, P. C. Wright, and P. Olivier, “A pool of dreams: Facebook, politics and the emergence of a social movement,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’14, Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 3573–3582, ISBN: 9781450324731.

[63] S. Savage, A. Monroy-Hernandez, and T. Höllerer, “Botivist: Calling volunteers to action using online bots,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ser. CSCW ’16, San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 813–822, ISBN: 9781450335928.

[64] L. Palen *et al.*, “A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters,” *ACM-BCS Visions of Computer Science 2010*, pp. 1–12, 2010.

[65] L. Palen and K. M. Anderson, “Crisis informatics—new data for extraordinary times,” *Science*, vol. 353, no. 6296, pp. 224–225, 2016. eprint: https://science.sciencemag.org/content/353/6296/224.full.pdf.

[66] R. Soden and L. Palen, “From crowdsourced mapping to community mapping: The post-earthquake work of openstreetmap haiti,” in *COOP 2014-Proceedings of the 11th International Conference on the Design of Cooperative Systems, 27-30 May 2014, Nice (France)*, Springer, 2014, pp. 311–326.

[67] R. Soden, L. Sprain, and L. Palen, “Thin grey lines: Confrontations with risk on colorado’s front range.,” in *CHI*, 2017, pp. 2042–2053.

[68] M. Zook, M. Graham, T. Shelton, and S. Gorman, “Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake,” *World Medical & Health Policy*, vol. 2, no. 2, pp. 7–33, 2010.

[69] J. An, H. Kwak, Y. Mejova, D. Oger, S. A. Saenz, and B. G. Fortes, “Are you charlie or ahmed? cultural pluralism in charlie hebdo response on twitter,” *arXiv preprint arXiv:1603.00646*, 2016.

[70] P. Burnap *et al.*, “Tweeting the terror: Modelling the social media reaction to the woolwich terrorist attack,” *Social Network Analysis and Mining*, vol. 4, no. 1, p. 206, 2014.

[71] S. Wiegand and S. E. Middleton, "Veracity and velocity of social media content during breaking news: Analysis of november 2015 paris shootings," in *Proceedings of the 25th international conference companion on world wide web*, 2016, pp. 751–756.

[72] I. C.-H. Fung, Z. T. H. Tse, C.-N. Cheung, A. S. Miu, and K.-W. Fu, "Ebola and the social media.," 2014.

[73] C. Reuter and M.-A. Kaufhold, "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics," *Journal of Contingencies and Crisis Management*, vol. 26, no. 1, pp. 41–57, 2018.

[74] R. Soden and L. Palen, "Informating crisis: Expanding critical perspectives in crisis informatics," *Proceedings of the ACM on human-computer interaction*, vol. 2, no. CSCW, pp. 1–22, 2018.

[75] P. Olivier and P. Wright, "Digital civics: Taking a local turn," *Interactions*, vol. 22, no. 4, pp. 61–63, Jun. 2015.

[76] S. Farnham, D. Keyes, V. Yuki, and C. Tugwell, "Puget sound off: Fostering youth civic engagement through citizen journalism," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, ser. CSCW '12, Seattle, Washington, USA: Association for Computing Machinery, 2012, pp. 285–294, ISBN: 9781450310864.

[77] J. E. Coronado Escobar and A. R. Vasquez Urriago, "Gamification: An effective mechanism to promote civic engagement and generate trust?" In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 2014, pp. 514–515.

[78] E. Gordon and G. Koo, "Placeworlds: Using virtual worlds to foster civic engagement," *Space and Culture*, vol. 11, no. 3, pp. 204–221, 2008.

[79] E. Gordon, S. Schirra, and J. Hollander, "Immersive planning: A conceptual model for designing public participation with new technologies," *Environment and Planning B: Planning and Design*, vol. 38, no. 3, pp. 505–519, 2011.

[80] C. DiSalvo, J. Lukens, T. Lodato, T. Jenkins, and T. Kim, "Making public things: How hci design can express matters of concern," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 2397–2406.

[81] S. Fox, R. R. Ulgado, and D. Rosner, "Hacking culture, not devices: Access and recognition in feminist hackerspaces," in *Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing*, 2015, pp. 56–68.

[82] S. Lindtner, G. D. Hertz, and P. Dourish, "Emerging sites of hci innovation: Hackerspaces, hardware startups & incubators," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 439–448.

[83] M. Balestrini, Y. Rogers, and P. Marshall, "Civically engaged hci: Tensions between novelty and social impact," in *Proceedings of the 2015 British HCI Conference*, ser. British HCI '15, Lincoln, Lincolnshire, United Kingdom: Association for Computing Machinery, 2015, pp. 35–36, ISBN: 9781450336437.

[84] M. Asad, C. A. Le Dantec, B. Nielsen, and K. Diedrick, "Creating a sociotechnical api: Designing city-scale community engagement," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17, Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2295–2306, ISBN: 9781450346559.

[85] E. Corbett and C. A. Le Dantec, "Exploring trust in digital civics," in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS '18, Hong Kong, China: Association for Computing Machinery, 2018, pp. 9–20, ISBN: 9781450351980.

[86] C. Crivellaro, R. Comber, M. Dade-Robertson, S. J. Bowen, P. C. Wright, and P. Olivier, "Contesting the city: Enacting the political through digitally supported urban walks," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15, Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 2853–2862, ISBN: 9781450331456.

[87] C. DiSalvo and C. A. L. Dantec, "Civic design," *Interactions*, vol. 24, no. 6, pp. 66–69, Oct. 2017.

[88] M. Harding, B. Knowles, N. Davies, and M. Rouncefield, "Hci, civic engagement, and trust," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15, Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 2833–2842, ISBN: 9781450331456.

[89] N. Moreno, S. Savage, A. Leal, J. Cornick, M. Turk, and T. Höllerer, "Motivating crowds to volunteer neighborhood data," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and Social Computing*, ser. CSCW'15 Companion, Vancouver, BC, Canada: Association for Computing Machinery, 2015, pp. 235–238, ISBN: 9781450329460.

[90] A. Bowser *et al.*, "Using gamification to inspire new citizen science volunteers," in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, ser. Gamification '13, Toronto, Ontario, Canada: Association for Computing Machinery, 2013, pp. 18–25, ISBN: 9781450328159.

[91] E. Massung, D. Coyle, K. F. Cater, M. Jay, and C. Preist, "Using crowdsourcing to support pro-environmental community activism," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13, Paris, France: Association for Computing Machinery, 2013, pp. 371–380, ISBN: 9781450318990.

[92] C. A. Le Dantec, M. Asad, A. Misra, and K. E. Watkins, "Planning with crowdsourced data: Rhetoric and representation in transportation planning," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15, Vancouver, BC, Canada: Association for Computing Machinery, 2015, pp. 1717–1727, ISBN: 9781450329224.

[93] C. A. Le Dantec, *Designing publics*. MIT Press, 2016.

[94] S. Abdullah, E. L. Murnane, J. M. Costa, and T. Choudhury, "Collective smile: Measuring societal happiness from geolocated images," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work amp; Social Computing*, ser. CSCW '15, Vancouver, BC, Canada: Association for Computing Machinery, 2015, pp. 361–374, ISBN: 9781450329224.

[95] J. Chen, G. Hsieh, J. U. Mahmud, and J. Nichols, "Understanding individuals' personal values from social media word use," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work amp; Social Computing*, ser. CSCW '14, Baltimore, Maryland, USA: Association for Computing Machinery, 2014, pp. 405–414, ISBN: 9781450325400.

[96] V. R. K. Garimella, A. Alfayad, and I. Weber, "Social media image analysis for public health," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, San Jose, California, USA: Association for Computing Machinery, 2016, pp. 5543–5547, ISBN: 9781450333627.

[97] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "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*, ser. CHI '16, San Jose, California, USA: Association for Computing Machinery, 2016, pp. 2098–2110, ISBN: 9781450333627.

[98] S. Bagroy, P. Kumaraguru, and M. De Choudhury, "A social media based index of mental well-being in college campuses," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 1634–1646, ISBN: 9781450346559.

[99] M. De Choudhury, S. S. Sharma, T. Logar, W. Eekhout, and R. C. Nielsen, "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, ser. CSCW '17, Portland, Oregon, USA: Association for Computing Machinery, 2017, pp. 353–369, ISBN: 9781450343350.

[100] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2013, pp. 3267–3276.

[101] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," ser. CSCW '14, Baltimore, Maryland, USA: Association for Computing Machinery, 2014, pp. 626–638, ISBN: 9781450325400.

[102] K. Saha and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, Dec. 2017.

[103] C. Homan, R. Johar, T. Liu, M. Lytle, V. Silenzio, and C. O. Alm, "Toward macro-insights for suicide prevention: Analyzing fine-grained distress at scale," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, pp. 107–117.

[104] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from twitter activity," ser. CHI '15, Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 3187–3196, ISBN: 9781450331456.

[105] P. Kostkova, "Public health," in *Twitter: A Digital Socioscope*, Y. Mejova, I. Weber, and M. W. Macy, Eds. Cambridge University Press, 2015, pp. 111–130.

[106] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson, "Identifying health-related topics on twitter," in *Social Computing, Behavioral-Cultural Modeling and Prediction*, J. Salerno, S. J. Yang, D. Nau, and S.-K. Chai, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 18–25, ISBN: 978-3-642-19656-0.

[107] J. Parker, Y. Wei, A. Yates, O. Frieder, and N. Goharian, "A framework for detecting public health trends with twitter," in *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, 2013, pp. 556–563.

[108] M. Paul and M. Dredze, "You are what you tweet: Analyzing twitter for public health," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, pp. 265–272, Aug. 2021.

[109] S. Chancellor, T. Mitra, and M. De Choudhury, "Recovery amid pro-anorexia: Analysis of recovery in social media," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 2111–2123.

[110] D. Won, Z. C. Steinert-Threlkeld, and J. Joo, "Protest activity detection and perceived violence estimation from social media images," *CoRR*, vol. abs/1709.06204, 2017. arXiv: 1709.06204.

[111] V. Marivate and P. Moiloa, "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, 2016, pp. 1–5.

[112] M. S. Gerber, "Predicting crime using twitter and kernel density estimation," *Decision Support Systems*, vol. 61, pp. 115–125, 2014.

[113] S. K. Rumi, K. Deng, and F. D. Salim, "Crime event prediction with dynamic features," *EPJ Data Science*, vol. 7, no. 1, p. 43, 2018.

[114] C. Featherstone, "Identifying vehicle descriptions in microblogging text with the aim of reducing or predicting crime," in *2013 International Conference on Adaptive Science and Technology*, 2013, pp. 1–8.

[115] S. Aghababaei and M. Makrehchi, "Mining social media content for crime prediction," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, IEEE, 2016, pp. 526–531.

[116] X. Chen, Y. Cho, and S. Y. Jang, "Crime prediction using twitter sentiment and weather," in *2015 Systems and Information Engineering Design Symposium*, IEEE, 2015, pp. 63–68.

[117] N. Malleson and M. A. Andresen, "Spatio-temporal crime hotspots and the ambient population," *Crime science*, vol. 4, no. 1, pp. 1–8, 2015.

[118] J. Vertesi and P. Dourish, "The value of data: Considering the context of production in data economies," in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ser. CSCW '11, Hangzhou, China: Association for Computing Machinery, 2011, pp. 533–542, ISBN: 9781450305563.

[119] M. Feinberg, "A design perspective on data," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17, Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2952–2963, ISBN: 9781450346559.

[120] J. Lave and E. Wenger, *Situated learning: legitimate peripheral participation*. Cambridge University Press, 2018.

[121] J. Bates, Y.-W. Lin, and P. Goodale, “Data journeys: Capturing the socio-material constitution of data objects and flows,” *Big Data & Society*, vol. 3, no. 2, p. 2 053 951 716 654 502, 2016. eprint: https://doi.org/10.1177/2053951716654502.

[122] L. Daston and P. Galison, *Objectivity*. Zone Books, 2018.

[123] G. Latzko-Toth, C. Bonneau, and M. Millette, *Chapter 13: Small data, thick data: Thickening strategies for trace-based social media research*.

[124] J. Vitak, K. Shilton, and Z. Ashktorab, “Beyond the belmont principles: Ethical challenges, practices, and beliefs in the online data research community,” in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 941–953.

[125] E. Horvitz and D. Mulligan, “Data, privacy, and the greater good,” *Science*, vol. 349, no. 6245, pp. 253–255, 2015. eprint: https://science.sciencemag.org/content/349/6245/253.full.pdf.

[126] J. Metcalf and K. Crawford, “Where are human subjects in big data research? the emerging ethics divide,” *Big Data & Society*, vol. 3, no. 1, p. 2 053 951 716 650 211, 2016.

[127] C. Fiesler, N. Beard, and B. C. Keegan, “No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 187–196.

[128] E. Department of Health *et al.*, “The belmont report. ethical principles and guidelines for the protection of human subjects of research.,” *The Journal of the American College of Dentists*, vol. 81, no. 3, p. 4, 2014.

[129] J. Lingel, *Digital Countercultures and the Struggle for Community*. Apr. 2017, ISBN: 9780262036214.

[130] C. A. Le Dantec, C. Appleton, M. Asad, R. Rosenberger, and K. Watkins, “Advocating through data: Community visibilities in crowdsourced cycling data,” in *Bicycle Justice and Urban Transformation*, 1st. Routledge, 2016.

[131] S. Lewis and D. A. Lewis, “Examining technology that supports community policing,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’12, Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 1371–1380, ISBN: 9781450310154.

[132] N. Sachdeva and P. Kumaraguru, “Online social media and police in india: Behavior, perceptions, challenges,” *CoRR*, vol. abs/1403.2042, 2014. arXiv: 1403.2042.

[133] P. Alston, "Introduction: Third generation human rights fact-finding," *Proceedings of the ASIL Annual Meeting*, vol. 107, pp. 61–62, 2013.

[134] J. D. Aronson, "The utility of user-generated content in human rights investigations," in *New Technologies for Human Rights Law and Practice*, M. K. Land and J. D. Aronson, Eds. Cambridge University Press, 2018, pp. 129–148.

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

[136] M. Browne, *Youtube removes videos showing atrocities in syria*, Aug. 2017.

[137] N. Cumming-bruce, *Ex-judge chosen by u.n. to gather evidence of syria war crimes*, Jul. 2017.

[138] K. McEvers, *Youtube inadvertently erases syrian war videos in purge of extremist propaganda*, Sep. 2017.

[139] A. International, *Deadly detention:deaths in custody amid popular protest in syria*, 2011.

[140] ——, *How is social media transforming human rights monitoring?* Feb. 2013.

[141] E. McPherson, "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.

[142] C. Reuter and M.-A. Kaufhold, "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics," *Journal of Contingencies and Crisis Management*, vol. 26, no. 1, pp. 41–57, 2018.

[143] C. M. de Defensa y Promoción de los Derechos Humanos, "Violaciones graves a derechos humanos en la guerra contra las drogas en méxico," *Comisión Mexicana de Defensa y Promoción de los Derechos Humanos, a.c.*, Mar. 2015.

[144] *Resolution on disappearance of people in mexico*, Nov. 2019.

[145] *Un rights chief urges action in mexico to end 'outrageous' wave of disappearances — un news*, May 2018.

[146] *Mexico: Damning un report highlights cover-up in case of 43 disappeared students*, Mar. 2018.

[147] Wola, *Mexico must summon political will to fully implement law on forced disappearances*, Aug. 2018.

[148] D. Franco-Migues, “La opacidad en ”el registro” de desaparecidos en méxico,” *Del dicho al hecho: opacidad, autoritarismo y verdades a medias*, Apr. 2015.

[149] P. Ferri, *”después de contar otra vez, la cifra de desaparecidos en méxico va a aumentar un 30%”*, Dec. 2019.

[150] C. G. Mena, *Onu: Tragedia humanitaria en méxico por desapariciones*, Aug. 2019.

[151] Forbes, *México no tiene una cifra oficial de desaparecidos • forbes méxico*, Aug. 2019.

[152] U. I. C. D. México and A. Comisión Mexicana de Defensa y Promoción de Derechos Humanos, *Violencia y terror: Hallazgos sobre fosas clandestinas en méxico*, 2018.

[153] M. Meltis, J. Ruiz, and P. Ball, *New results for the identification of municipalities with clandestine graves in mexico*, Nov. 2017.

[154] C. Grillo, *Locating hidden graves in mexico*, Mar. 2017.

[155] P. Villegas, *168 skulls in mexican mass grave are reminder of cartel violence*, Sep. 2018.

[156] L. Wade, *Mapping mexico’s hidden graves*, Jun. 2017.

[157] X. Antillon, *El movimiento por nuestros desaparecidos en méxico*, Nov. 2015.

[158] C. de Colaboración Cívica and M. Foundation, *El movimiento por nuestros desaparecidos en méxico y su camino hacia la incidencia legislativa: La siembra colectiva, una apuesta por la esperanza*, Aug. 2018.

[159] J. Montes and J. de Córdoba, *’it’s a crisis of civilization in mexico.’ 250,000 dead. 37,400 missing.* Nov. 2018.

[160] F. U. por Nuestros Desaparecidos en Nuevo León, *Un sentido de vida: La experiencia de búsqueda de fuerzas unidas por nuestros desaparecidos en nuevo león 2012–2019*, Jan. 2019.

[161] I. López, *Familiares exigen participar en la creación de la ley de desaparición forzada en la cdmx*, Oct. 2019.

[162] *El movimiento por nuestros desaparecidos en méxico trabaja para que la #leydesaparición se implemente correctamente en todo méxico*, 2018.

[163] *Onu-dh: Ley de desaparición debe pasar de lo formal a lo real*, Jan. 2019.

[164] INEGI, *Encuesta nacional de seguridad pública urbana (ensu) tercer trimestre 2019*, Mar. 2018.

[165] ——, *Encuesta nacional de seguridad pública urbana (ensu) tercer trimestre 2019*, Oct. 2019.

[166] *Base de datos ante mortem/post mortem: Herramienta informática para la gestión de datos forenses*, 2014.

[167] *Impunidad estadística en méxico: Los delitos que ignoramos*, Mar. 2017.

[168] B. Schneiderman *et al.*, *Nodexl: A free and open network overview, discovery and exploration add-in for excel*, 2010.

[169] M. Santiago Juárez, J. C. Gutiérrez Contreras, and Z. Valenzuela Contreras, *Guía práctica sobre la aplicación del protocolo homologado para la búsqueda de personas desaparecidas*, Apr. 2017.

[170] H. R. Watch, *Mexico: Lessons from a human rights catastrophe*, Jan. 2019.

[171] D. F. Orentlicher, *Bearing witness: The art and science of human rights fact-finding*, 1990.

[172] T. Gillespie, "Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us," *Culture Digitally*, 2011.

[173] T. Gillespie, "The politics of 'platforms'," *new media & society*, vol. 12, no. 3, pp. 347–364, May 2010.

[174] M. Madianou, "Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises," *Social Media + Society*, vol. 5, no. 3, p. 2 056 305 119 863 146, 2019.

[175] Z. Tufekci, "Big questions for social media big data: Representativeness, validity and other methodological pitfalls," *ArXiv*, vol. abs/1403.7400, 2014.

[176] A. Olteanu, C. Castillo, F. Diaz, and E. Kıciman, "Social data: Biases, methodological pitfalls, and ethical boundaries," *Frontiers in Big Data*, vol. 2, p. 13, 2019.

[177] N. Baym, "Data not seen: The uses and shortcomings of social media metrics," *First Monday*, vol. 18, 2013.

[178] L. McCay-Peet and A. Quan-Haase, "What is social media and what questions can social media research help us answer," 2016.

[179] M. Kogan, A. Halfaker, S. Guha, C. Aragon, M. Muller, and S. Geiger, "Mapping out human-centered data science: Methods, approaches, and best practices," in *Companion of the 2020 ACM International Conference on Supporting Group Work*, ser. GROUP '20, Sanibel Island, Florida, USA: Association for Computing Machinery, 2020, pp. 151–156, ISBN: 9781450367677.

[180] M. Muller *et al.*, "Human-centered study of data science work practices," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '19, Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–8, ISBN: 9781450359719.

[181] *Nuevo coronavirus 2019*.

[182] M. Digital, Nov. 2020.

[183] *Coronavirus en méxico. confirman primera muerte por covid-19*, Mar. 2020.

[184] *Semáforo*.

[185] COVID-19, *Se mantienen semáforo rojo y suspensión de actividades no esenciales en la ciudad de méxico*, Jan. 2021.

[186] C. Alvino, *Estadísticas de la situación digital de méxico en el 2020-2021*, Jun. 2021.

[187] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Workshop Proceedings of the International Conference on Learning Representations*, ser. ICLR '13, Scottsdale, Arizona, USA, 2013.

[188] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, pp. 3111–3119, 2013.

[189] C. Cardellino, *Spanish Billion Words Corpus and Embeddings*, Aug. 2019.

[190] C. T. d. l. M. Quiñonez Amezquita, Ph.D. dissertation, 2005.

[191] M. Muller *et al.*, "Designing ground truth and the social life of labels," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021, ISBN: 9781450380966.

[192] L. Aroyo and C. Welty, "Truth is a lie: Crowd truth and the seven myths of human annotation," *AI Magazine*, vol. 36, no. 1, pp. 15–24, Mar. 2015.

[193] M. Muller *et al.*, "Interrogating data science," in *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '20 Companion, Virtual Event, USA: Association for Computing Machinery, 2020, pp. 467–473, ISBN: 9781450380591.

[194] A. X. Zhang, M. Muller, and D. Wang, "How do data science workers collaborate? roles, workflows, and tools," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW1, May 2020.

[195] T. M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press, 2020, ISBN: $10.1515_9780691210544$.

[196] S. Greenberg, "Toolkits and interface creativity," *Multimedia Tools and Applications*, vol. 32, no. 2, pp. 139–159, 2006.

[197] D. Ledo, S. Houben, J. Vermeulen, N. Marquardt, L. Oehlberg, and S. Greenberg, "Evaluation strategies for hci toolkit research," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18, Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–17, ISBN: 9781450356206.

[198] N. Marquardt, S. Houben, M. Beaudouin-Lafon, and A. D. Wilson, "Hcitools: Strategies and best practices for designing, evaluating and sharing technical hci toolkits," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17, Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 624–627, ISBN: 9781450346566.

[199] C. Holt, S. Fawcett, V. Francisco, J. Schultz, B. Berkowitz, and T. Wolff, *The community tool box*.

[200] T. Keane *et al.*, *Diy development impact and you: Practical tools to trigger and support social innovation*.

[201] B. Evaluation, *Rainbow framework*, Aug. 2014.

[202] A. International, *Toolkits and guides*.

[203] F. Pfortmüller, N. Luchsinger, and S. Mombartz, *The community canvas*, May 2017.

[204] S. Mattern, *Unboxing the toolkit*, Jul. 2021.

[205] R. Y. Wong, M. A. Madaio, and N. Merrill, *Seeing like a toolkit: How toolkits envision the work of ai ethics*, 2022.

[206] Panel contributors., *Panel Documentation*, https://panel.holoviz.org, Accessed: 2022-04-11, 2022.

[207] Tweeter, Inc., *What's new with v2*, https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api, Accessed: 2022-04-11, 2022.

[208] Joshua Roesslein, *Tweepy Documentation*, https://docs.tweepy.org/en/stable/index.html, Accessed: 2022-04-11, 2022.

[209] Tweeter, Inc., *Building queries for Search Tweets*, https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query, Accessed: 2022-04-11, 2022.

[210] C. Cardellino, *Spanish Billion Words Corpus and Embeddings*, https://crscardellino.github.io/SBWCE/, Accessed: 2022-04-11, 2019.

[211] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, http://is.muni.cz/publication/884893/en, Valletta, Malta: ELRA, May 2010, pp. 45–50.

[212] D. Harawy, "Situated knowledges: The science question in feminism and the privilege of partial perspective," *Feminist Studies*, vol. 14, no. 3, pp. 575–599, 1988.