# Data Migrations: Exploring the Use of Social Media Data as Evidence for Human Rights Advocacy

ADRIANA ALVARADO GARCIA, Georgia Institute of Technology
MATTHEW J. BRITTON, Muso Health
DHAIRYA MANISH DOSHI, Amazon
MUNMUN DE CHOUDHURY, Georgia Institute of Technology
CHRISTOPHER A. LE DANTEC, Georgia Institute of Technology

Social media platforms offer a rich repository of crowdsourced information that has the potential to monitor human rights violations. The challenge is to quantify, interpret, and situate such unstructured data streams in the broader context, which remains under-investigated in existing CSCW research. Addressing these challenges demands computational solutions to extract large volumes of data in conjunction with human intervention to transition the data streams into the *offline* context to render them usable and actionable. Following an iterative human-in-the-loop computational approach, we explore whether citizen reports of abductions concentrated on Facebook groups can be useful to complete official records on the ongoing crisis of disappearances in Mexico. We conceptualize three key practices of the process of transitioning the data from online to offline, followed by seven qualitative characteristics of the data streams that contribute to each stage of the process. Our research contributes with an initial understanding of the challenges and opportunities of migrating the local knowledge from online communities to be used as evidence by organizations seeking to address institutional failures.

CCS Concepts: • **Information systems → Collaborative and social computing systems and tools**; • **Human-centered computing** → *Collaborative and social computing systems and tools.*

Additional Key Words and Phrases: data practices, data migration, human rights, missing people, Mexico, social media platforms, Facebook, activism, social justice

**ACM Reference Format:**
Adriana Alvarado Garcia, Matthew J. Britton, Dhairya Manish Doshi, Munmun De Choudhury, and Christopher A. Le Dantec. 2020. Data Migrations: Exploring the Use of Social Media Data as Evidence for Human Rights Advocacy. *J. ACM* 37, 4, Article 111 (August 2020), 25 pages. https://doi.org/10.1145/1122445.1122456

## 1 INTRODUCTION

Social media sites have established themselves as more than ludic outlets for individual engagement and have become platforms that enable citizens to organize through the accumulation of crowd-sourced information. We know from prior work that people turn to these platforms in moments of crisis [33, 79], when in search of emotional support [15, 21], and to take up volunteer efforts to address local issues [30, 89]. As these diverse and serious uses of platforms like Facebook, Twitter,

Authors' addresses: Adriana Alvarado Garcia, adriana.ag@gatech.edu Georgia Institute of Technology; Matthew J. Britton, matthew.james.britton@gmail.com Muso Health; Dhairya Manish Doshi, dhairyadoshi96@gmail.com Amazon; Munmun De Choudhury, munmun.choudhury@cc.gatech.edu Georgia Institute of Technology; Christopher A. Le Dantec, ledantec@gatech.edu Georgia Institute of Technology.


0004-5411/2020/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

and Instagram have been taken hold, we can recognize the rich opportunities to bridge civic interactions where governments and citizens alike use the affordances of social media to inform [72], organize [88, 96], and take action on local issues [31, 39].

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 [12]. Examples include tracing incidents of mental health crises on college campuses [85], or, in our case, tracking an ongoing international human rights crises. Building on exiting work, we know that social media platforms can be turned to for evidence to hold institutions accountable [18, 19, 93]. An example of this practice are the hundreds of YouTube videos documenting human rights violations in the civil war in Syria [26, 32, 68]. Similarly, social media data has been useful in identifying cases of violence against specific populations in the absence of official records [58, 59], and in extending the categories of data collected by governments during humanitarian crises [69, 83]. 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 (NGOs) 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 [18, 34]. We 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 [27, 98]. Among the most prominent challenges are: how to make data legible to different stakeholders and "rendered transportable and intelligible" [27], 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" [27], 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 [27].

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. Our research 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 the work we are presenting, we examine reports of missing persons in Mexico that were shared via Facebook groups. Through an iterative process, we identified seven data characteristics that helped us to migrate 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 circulating on Facebook. We 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 [38, 45, 97]. 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 [7, 8, 10, 108]. 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 [43, 44, 46, 71, 74]. The situation has only become more urgent as recent discoveries of clandestine mass graves with unidentified bodies throughout the country have come to light [52, 70, 74, 99, 100].

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 [16, 37]. One of the most notable is the collective called *Movement for our disappeared in Mexico*[1], which is composed of more than 60 NGOs from 21 states. These NGOs are responsible for finding a large number of clandestine mass graves [73, 74, 80], 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 [6, 9, 64, 80]. 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 [37].

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 [13]. Our contribution is to first operationalize the migration of social media data to an offline context. We make this contribution by developing three stages of the migration process and the qualitative characteristics of the data stream that render this migration possible. Second, by establishing a data gap in the official database of missing people, we depict how data migration and the characteristics we identified can be used to support evidence gathering by NGOs. In this context, and building on prior work, we define evidence as to any material or data that meet any of the following criteria: 1) provides a starting point for investigation, 2) signifies the occurrence of a crime, 3) links a specific person or group to the violation of a human right, 4) shows a pattern of behavior that may contribute to combating state responses to allegations of abuse, 5) helps to determine what happened at a particular time and place when an event happened, or 6) establishes the conditions in a place that is no longer accessible by police or human rights investigators [17, 81].

In conducting this research, we aim to provide insight into a class of problems where the site of data use is divorced from the site of data collection or production. By developing a data migration process and the data attributes that support this procedure, we are working to transition from individual responses to human-rights violations to collective and systemic capacities by building a bridge between community data practices and institutions such as NGOs. As more communities turn to computing infrastructures to organize and work towards social change, the need to design and support locally informed systems of knowledge production only becomes more urgent.

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

## 2 THE DRUG WAR IN MEXICO

Mexico is experiencing a renewed crisis of human rights violations and extreme violence coordinated by organized crime as well as state security forces [45, 94, 102, 105]. One of the main characteristics of this crisis is the re-emergence of enforced disappearances as a widespread problem [38, 102]. As described by the law establishing the National Registry of Missing or Disappeared Persons, a disappeared person is "Any person whose whereabouts are unknown to their family or who, based on reliable information, has been declared absent according with domestic law concerning an international or non-international armed conflict, a situation of violence or disturbances, a natural catastrophe or any situation that may require the intervention of a competent public authority" [5]. In vernacular terms, to "disappear" in Mexico is not only to be forcibly abducted, but also to be ignored and forgotten by the government [28, 37, 82, 97, 103, 106].

The roots of the current crisis can be traced back to the 1960's during the Dirty War (Guerra Sucia), when Mexican security forces carried out the systematic repression of anyone suspected of being part of the opposition movement [28, 82, 106]. During the Dirty War, there were no legal protections to help victims' families search for missing family members or investigate reported abductions [28, 82]. The Dirty War ended in the 1980's but the exact number of victims has still not been officially recognized despite several different government commissions that were created to clarify what happened and find those responsible [82]. It is within this larger historic context that enforced disappearances have re-emerged via the War on Drugs that started in 2006. As a result, by 2011, the Mexican government enacted the General Law on Forced Disappearance of Persons recognizing the crisis and outlining a strategy to address the root problem [97]. Despite the implementation of this legislation and the creation of multiple specialized government agencies and databases, victims' families and NGOs continue to bear the burden of documenting and searching for missing persons [16, 47, 80].

The human rights crisis that Mexico is currently experiencing is, in part, a consequence of the fact that violence and disappearance that took place during the Dirty War were never addressed by the government [106]. The disappearance of people has been an uninterrupted practice since the 1960's and the government keeps failing to provide an accurate figure on the number of missing people, as well as providing an explanation for those disappearances [53, 75, 106]. As a response to the crises, a large number of NGOs have emerged. Some remarkable examples include the work conducted by *United Forces for Our Disappeared in Nuevo León* [2] and *The Others Missing* [3]. These NGOs are part of the collective *Movement for our disappeared in Mexico*, and offer legal and psychological support to the relatives of missing people, and organize search squads to recover bodies found in clandestine graves.

## 3 RELATED WORK

Addressing structural social ills like crime and violence require first understanding the magnitude of the problem and then building policy and legal responses that intervene at the appropriate scale – local, state, nation. Computing, from automation to data collection, plays an important role in both gaining insight into the conditions and locations of crime and in supporting government and non-government entities as they seek to support victims and develop structural remedies.

[2] In Spanish: Fuerzas Unidas por Nuestros Desaparecidos en Nuevo León (FUNDENL). This is an organization initiated by relatives of missing people from the state of Nuevo León (http://fundenl.org/)

[3] In Spanish: Los Otros Desaparecidos de Iguala, Guerrero. This is an NGO in operation since 2014 that works with families of missing persons in the state of Guerrero, Mexico (http://losotrosdesaparecidos.org/)

### 3.1 Crime Detection and Prediction with Social Media

Like other crises, the emergence of social media and crowdsourced data create a new, distributed, and accessible set of tools for the public to directly catalogue and share information about the causes and consequences of enforced disappearance. The use of Information and Communication Technologies (ICTs) in supporting citizen activism [39, 48], citizen participation in crime prevention [41, 62], and the design of crime prevention tools has long been an area of focus for CSCW and HCI more broadly [23, 60, 87]. More recently, researchers have examined the role of ICTs in armed conflict and how ICTs support local responses to the unique challenges of recovering from widespread violence and destruction [35, 55, 92]. By drawing together research from computational crime prediction we can begin to layout the landscape for how to use social media data to confront the current crisis of human rights violations in Mexico.

A growing body of computational research has begun to focus on assessing 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 [109],leveraging Twitter data to better understand patterns in incidents of crime and creating classification models to automatically label incoming social media data [67]; using Twitter data drawn from local news agencies to predict hit-and-run vehicular accidents and breaking-and-entering crimes [101]; 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 [49, 84]. Additional work has used Twitter data to augment and assess vehicular descriptions linked to crime [42]; forecasting likely changes in crime over time using a number of different social and environmental factors [11, 29, 107]; 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 [66].

While prior work has been of interest to criminologists [61, 95], criminal justice advocates [54], and law enforcement organizations [40], much of it focuses on data from public criminal records and data collected specifically from Twitter. Such an approach works well in social and cultural contexts where data tracked by law enforcement can be treated as "ground truth" and where Twitter is in widespread use – like the in the U.S. However, when looking to other geopolitical contexts, the realities around both the dearth of robust crime data, and different preferences for social media platforms raise the need to re-examine those assumptions and to use local constraints to both shape how official data are treated and which social media data streams provide meaningful signals for tracking and understanding public safety, and at the extreme, human-rights violations.

### 3.2 Technology and Human Rights Advocacy

Turning to the conditions in Mexico and the larger milieu of supporting human rights globally, a small but important number of studies have begun to emerge in HCI. These have ranged from investigating the emergence of "curators" – individuals who act as "war correspondents" – who aggregate and disseminate information to large social media audiences [72]; to focusing on the affective impact violent crime has had on local residents [36]; and to understanding how armed groups of civilians known as "self-defense forces" have emerged through online channels to challenge the powerful drug cartels [88]. In the last instance, these groups have adopted pro-militia propagandist roles and online mobilization strategies to define the narrative of the armed conflict [88]. In light of this work, although Mexico's war on drugs has not been officially declared an armed conflict zone, the conditions of violence and trauma in the communities we examined have similar challenges to those in recognized conflict zones including a dependency on the available (and inadequate) infrastructure and asymmetric access to ICTs [35, 92].

For human rights NGOs addressing these conditions, the core of their work is to use national and international legal systems to hold perpetrators accountable for human rights violations [69, 77]. A critical first step in these interventions involves gathering and evaluating information that may be used as evidence and then developing legal and policy strategies based on that evidence [77]. Just as new data sources and analysis techniques are changing how we are approaching crime prediction and prevention, ICTs are changing how human rights organizations documenting and conducting their advocacy work, including who does the work of documenting and advocating [78]. The expansion of ICTs – the internet, smartphones, social media platforms, mapping, and geospatial technologies – has become a powerful and ubiquitous tool facilitating the process of documenting human rights violations [78].

While ICTs facilitate the production, storage, and distribution of digital information, they also introduce new challenges for human rights defenders. Chief among these challenges is how to manage large new data sources – both the volume and verification of these data – for use as evidence in advocacy work [69]. The volume challenge relies on identifying relevant information to a particular incident and whether or not the information documents a violation. The verification challenge addresses the quality of information and whether that information is at sufficient fidelity to include it as a verified addition to the overall picture. Verification techniques demand a combination of digital tools, external databases, and human expertise to identify the source, time, place, and accuracy of information [69].

One response to these challenges has been the creation of various tools to reduce both the amount of time to identify relevant information and the expertise needed to evaluate it. Tools like the CameraV App and eyeWitness allow users to take photos and videos, and to automatically embed metadata so the content can be further analyzed [3, 20]. Additional tools address the verification challenges by harnessing the power of the crowd – systems like Veri.ly, a web-based tool that verifies visual and textual information via crowdsourcing, distributes human-verification and fact checking during humanitarian crises [2].

Although these specialized tools facilitate the collection of information by anyone, they are typically only used by human rights NGOs and fail to gain the power of scale that would make them more potent tools for data collection. On the other hand, social media platforms have become an established and important mechanism for activists and civilians to gather information and organize in the face of political and institutional collapse [96]. Due to their unique affordances, like contextualizing content through social ties and local narratives and making information flows personal and relatable [91, 96], social media platforms have the potential to become a tool that works at scale to gather evidence useful for addressing human rights violations [22, 34].

The existing body of research demonstrates the potential of social media as a data source for assessing and predicting crime. However, this prior research has been conducted in a narrow context of both geopolitical conditions and social media use: the particularities and limitations of "ground truth" which rely on existing robust data sources, established metrics, and rich analytic tools along with specific platform use and assumptions do not neatly translate across social, cultural, and political contexts. The challenges of how this class of unstructured, crowdsourced, and voluntary data might be *migrated* to address gaps in official data sources, rather than *augmenting* those sources, remains an understudied area of empirical research. This shift from augmenting to contesting official accounts of crime means we are not concerned with prediction models as such, and instead are more concerned with developing techniques to use social media data as a robust proxy measure for under-reported events. The question of how such proxy data can be *migrated* and integrated with offline – and typically partial and provisional – data counterparts is paramount.

## 4 CONTEXT AND 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 we conducted over the past four years, we 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 [13]. 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. We 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 we followed 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 [14].

We blended three stages of qualitative methodologies and computational methods into a recursive pipeline that helped us establish the *context* of the data we 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 data-set derived from a social media data stream that is otherwise noisy and imprecise. We 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 our method and results, we 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.

We focus on information gathered on Facebook because it is the most commonly used social media platform for learning about crime in Mexico [56, 57]. According to the National Urban Public Security Survey (ENSU by its acronym in Spanish)[4], 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 [56]. The top two sources of information were televised news and word of mouth with family and friends [56]. 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 [56, 57]. 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 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

[4] 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.

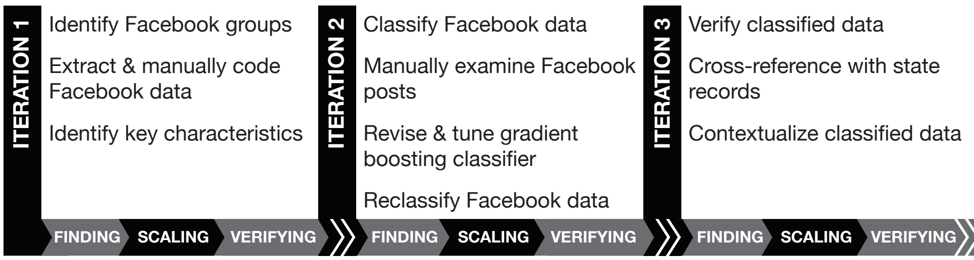


Fig. 1. Key practices we develop at each iteration

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.

### 4.1 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 [46, 97]. 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 [97]. Since each database has a different definition and approach for classifying who has gone missing, they all reflect a different number of victims [43, 46]. 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 Spanish)[5], which was the first official registry that coordinated across federal agency [46, 97]; 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) [1]. 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 [46]. 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 [4].

### 4.2 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 1). Each iteration comprised a mixed-methods approach where we

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

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, 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.

*4.2.1 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 [14]. 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 [13], 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 [90], 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 hand-labeled 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.

*4.2.2 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

[6]In Spanish: denuncia, vecinos, delegación.

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 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 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.

*4.2.3 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 [13]. 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

| 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<br>Verifying Signals |
| Redundant | Duplicate data | This attribute confirms the information and provided us with new insights. | Finding Signals<br>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 recov-ered 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<br>Verifying Signals |

Table 1. Definitions of the characteristics and the stages where they are used

gained at each iteration, and to further reflect on the interplay of qualitative and computational methods that enabled data migration.

## 5 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 we 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 [86].

Through our human-in-the-loop computational approach, we identified and confirmed 484 posts of missing people across the 45 Facebook pages we 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. Our 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 our analysis, we 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 1. The seven qualitative characteristics developed from each of the three iterations. This process was guided by our knowledge of the legislation

| 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 |

Table 2. Characteristics and Methods used at each stage

from Mexico and the techniques and methodological strategies NGOs have developed to identify the probative value of different types of evidence. Based on these elements, we looked for data characteristics that indicated a person went missing, the attributes that allowed us to maintain the context of the data as we migrated them, and how they contributed to the operational definition of evidence we outlined above. We describe each characteristic in the following section, linking them with the practices we consider necessary for migrating local knowledge from online communities into the offline context. Table 2 shows a concise description of the data characteristics and the methods relevant for each key practice of our process.

## 5.1 Finding Signals

In examining how to migrate social media data streams to the offline context of human rights advocacy, the first question we asked was *how do we identify signals in social media data streams that can be linked with official data records?* By asking this question, we aimed to identify whether or not there are social media data streams that document incidents of missing people.

During our qualitative analysis, we 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 2.

From the 484 posts that we identified, 333 included an image of a state-issued flier, while 151 included an image of a homemade flier. Distinguishing *official* posts helped us to identify data signaling the existence of an legitimate missing person claim, one that we would be able to verify


Fig. 2. Example of a state-issued flier of a missing person

in the national data registries of missing people (RNPED database). We 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.

### 5.2 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: we needed to establish the scale at which reports of missing people occur in social media data. Scaling signals helped us 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, we 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 the continuity of the Facebook communities. While conducting this research, one of the Facebook groups that we followed during the last three years, and from which we collected data, 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, we 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, we 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 we capture with the characteristic *situated*. Contrary to our expectations, our findings indicated that out of the 308 official posts, only 123 cases belonged to the municipalities covered by our 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 us to evaluate the scale at which the cases of missing people operate within social media, which was our second research question. These three attributes allow us to identify, quantify, and document cases of abductions from diverse online communities at specific times to then compare with the number of abductions reported in official databases.

### 5.3 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, we asked *how do we verify the*

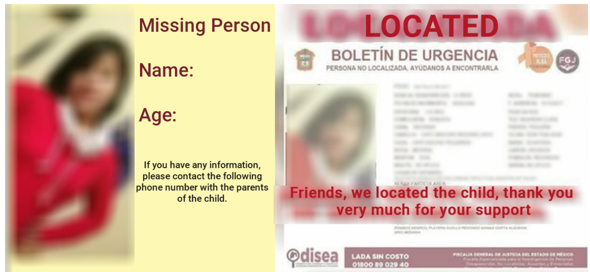


Fig. 3. Example of redundant information from an unofficial and a state-issued flier

*quality of social media data streams?*. We 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, we 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 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. We identified three primary sources of posts: those published 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 we were able to corroborate cases of missing people with reports in newspapers. In other cases, we 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 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, we use *legible* to describe the clarity of information encoded in state-issued fliers within the Facebook posts. As shown in Figure 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. We need the demographic information of each victim in order to search for them in the official database and make an accurate comparison between social media data streams and official data.

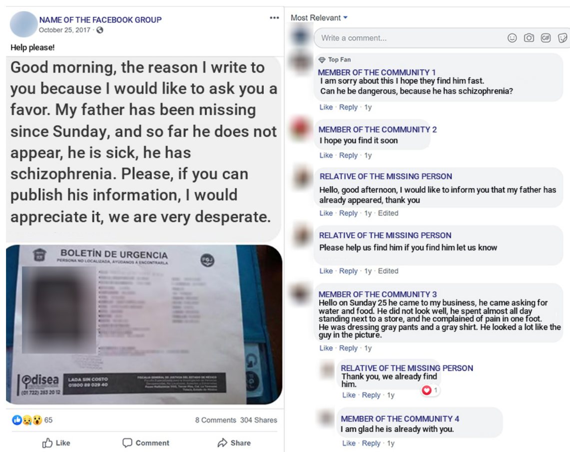


Fig. 4. Example of the descriptive attribute

From the 308 official posts, we 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). We 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). We 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 we only found 27 of 303 officially reported missing people confirms the inconsistencies among different sources about the number of missing people [43, 44, 46, 71, 74].

### 5.4 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 [97]. 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 [97].

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 [46]. 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 [97]).

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 [80].

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 [106]. In the second case, outside observers have noted that names had been consistently erased from national registries without offering an explanation to victims' relatives [104, 106].

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 [46, 97]. 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.

## 6 DISCUSSION

In working to understand how to develop social media data as a resource for NGOs developing evidence to confront human rights violations, we arrived at a set of practices that integrate computational and qualitative techniques for migrating data across contexts: finding signals, scaling signals, verifying signals. By applying these practices in an iterative manner, we were able to identify strong evidence of data gaps that exist in the official database of missing people in Mexico. While there remains additional work to further explain *why* those gaps exist, our contribution prioritizes the transferability of the characteristics that make *data migration* possible rather than a context-free generalizable tool. This last point is important and draws from critical data studies as well as design research where the context of data production is understood to be crucial and where the design of tools to solve problems comes in response to the particulars of problem, context, and capability [24, 27, 34, 76].

The data migration process we 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 [24, 25] – 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" [25]. Following Bowker's lead, we 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.

## 6.1 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, we 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* we 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, we were 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 we 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.

## 6.2 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 [17, 98]. As we 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 [14]. 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* – we combined computational and qualitative methods to properly identify and analyze trends and patterns of user-generated content. The former helped us tackle challenges of identifying and extracting volumes of data, while the latter methods guided our accounting of issues of representation and interpretation. At the *finding signals* stage, we 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 us 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 we 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 we developed, the data characteristics we 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 [63]. 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.

### 6.3 From Findings To Evidence

The seven data characteristics we identified helped us 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 our 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 [104, 106]. 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 [77]. 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 [12]. 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 [18]. 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 [17]. 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 [17].

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.

### 6.4 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 [13]. *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 [80]. 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 [96]. 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 [96]. 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 [50, 51, 96].

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 [65]. 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.

Lastly, as we move forward, we consider it useful to follow the guidelines suggested by Bowker to build appropriate tools for our goals [25]: when facing incompatible data structures among stakeholders, rather than creating unitary knowledge, advocate for plural representations that facilitate coordination. For it is the coordination across contexts – data contexts, platform contexts, institutional context – that is required to address threats to human rights.

## 7 CONCLUSION

In this paper, we examined the potential and limitations of using social media data streams to augment and contest official records of the ongoing crisis of disappearances in Mexico. Through a mixed methodology that interwove qualitative and computational methods, we were able to identify seven qualitative characteristics—*Official, Redundant, Descriptive, Retrievable, Quantifiable, Situated, and Legible*—of social media data streams that enable their migration to offline contexts for advocacy purposes. We then linked each qualitative characteristic to one of the three key data practices of developing social media data for use as offline evidence: finding the signal, understand the scale at which that signal is present, and verifying its accuracy. The outcome of our research provides an initial understanding of challenges and opportunities of migrating the local knowledge from online communities to be used as evidence by organizations seeking to address institutional failures. The empirical and methodological baseline we have established will enable future work to more specifically explore strategies that enable NGOs to take action with these new data and to support the particular and situated collaborative they do when confronting human rights violations in Mexico and other international contexts.

## 8 ACKNOWLEDGMENTS

Agradecemos a Marisol Wong - Villacrés y a nuestros revisores anónimos por sus valiosos comentarios. También queremos agradecer el apoyo de los miembros de la organización *Los Otros Desaparecidos,* cuyo trabajo continúa inspirándonos. Dedicamos este trabajo a todas aquellas personas en México y Latinoamérica que continúan en la búsqueda de sus seres queridos. This work was supported in part by the NSF under award IIS 1524380.

## REFERENCES

[1] 2014. Base De Datos Ante Mortem/Post Mortem: Herramienta Informática Para La Gestión De Datos Forenses. https://www.icrc.org/spa/assets/files/publications/icrc-003-4155.pdf

[2] 2014. Verily. https://veri.ly/
[3] 2015. CameraV: Secure Verifiable Photo and Video Camera. https://guardianproject.info/archive/camerav/
[4] 2017. Impunidad estadística en México: los delitos que ignoramos. https://seguridad.nexos.com.mx/?p=67
[5] 2017. Ley del Registro Nacional de Datos de Personas Extraviadas o Desaparecidas. http://www.diputados.gob.mx/LeyesBiblio/abro/lrndped/LRNDPED_abro.pdf
[6] 2018. El Movimiento por Nuestros Desaparecidos en México trabaja para que la #LeyDesaparición se implemente correctamente en todo México. https://movndmx.org/ley-desaparicion/
[7] 2018. Mexico: Damning UN report highlights cover-up in case of 43 disappeared students. https://www.amnesty.org/en/latest/news/2018/03/mexico-damning-un-report-highlights-cover-up-in-case-of-43-disappeared-students/
[8] 2018. UN rights chief urges action in Mexico to end 'outrageous' wave of disappearances | UN News. https://news.un.org/en/story/2018/05/1010981
[9] 2019. ONU-DH: Ley de Desaparición debe pasar de lo formal a lo real. https://www.hchr.org.mx/index.php?option=com_k2&view=item&id=1214:onu-dh-ley-de-desaparicion-debe-pasar-de-lo-formal-a-lo-real&Itemid=265
[10] 2019. Resolution on disappearance of people in Mexico. https://www.fidh.org/en/region/americas/mexico/resolution-on-disappearance-of-people-in-mexico
[11] Somayyeh Aghababaei and Masoud Makrehchi. 2016. Mining social media content for crime prediction. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 526–531.
[12] Philip Alston. 2013. Introduction: Third Generation Human Rights Fact-Finding. *Proceedings of the ASIL Annual Meeting* 107 (2013), 61–62. https://doi.org/10.5305/procannmeetasil.107.0061
[13] Adriana Alvarado Garcia and Christopher A. Le Dantec. 2018. Quotidian Report: Grassroots Data Practices to Address Public Safety. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 17 (Nov. 2018), 18 pages. https://doi.org/10.1145/3274286
[14] Adriana Alvarado Garcia, Alyson L. Young, and Lynn Dombrowski. 2017. On Making Data Actionable: How Activists Use Imperfect Data to Foster Social Change for Human Rights Violations in Mexico. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 19 (Dec. 2017), 19 pages. https://doi.org/10.1145/3134654
[15] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. 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 (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1485–1500. https://doi.org/10.1145/2998181.2998243
[16] Ximena Antillon. 2015. El Movimiento por Nuestros Desaparecidos en México. https://fundar.org.mx/el-movimiento-por-nuestros-desaparecidos-en-mexico/
[17] Jay Aronson. 2018. Challenging Impunity with User Generated Content. https://www.ijmonitor.org/2015/06/challenging-impunity-with-user-generated-content/
[18] Jay D. Aronson. 2014. Crowd-Training Machine Learning Systems for Human Rights Abuse Documentation. In *AAAI 2014*.
[19] Jay D. Aronson. 2018. *The Utility of User-Generated Content in Human Rights Investigations*. Cambridge University Press, 129–148. https://doi.org/10.1017/9781316838952.006
[20] International Bar Association. 2015. eyewitness. https://www.eyewitnessproject.org/
[21] Natalya N. Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. 2015. Social Sharing of Emotions on Facebook: Channel Differences, Satisfaction, and Replies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 154–164. https://doi.org/10.1145/2675133.2675297
[22] Fikret Berkes and Helen Ross. 2013. Community Resilience: Toward an Integrated Approach. *Society & Natural Resources* 26, 1 (2013), 5–20. https://doi.org/10.1080/08941920.2012.736605 arXiv:https://doi.org/10.1080/08941920.2012.736605
[23] Jan Blom, Divya Viswanathan, Mirjana Spasojevic, Janet Go, Karthik Acharya, and Robert Ahonius. 2010. Fear and the City: Role of Mobile Services in Harnessing Safety and Security in Urban Use Contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1841–1850. https://doi.org/10.1145/1753326.1753602
[24] Christine L. Borgman. 2016. *Big data, little data, no data: scholarship in the networked world.* The MIT Press.
[25] Geoffrey C. Bowker and Susan Leigh. Star. 2008. *Sorting things out: classification and its consequences.* MIT Press.
[26] Malachy Browne. 2017. YouTube Removes Videos Showing Atrocities in Syria. https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html
[27] Samuelle Carlson and Ben Anderson. 2007. What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication* 12, 2 (01 2007), 635–651. https://doi.org/10.1111/j.1083-6101.2007.00342.x arXiv:https://academic.oup.com/jcmc/article-pdf/12/2/635/22317230/jjcmcom0635.pdf
[28] Laura Castellanos. 2007. *México armado. 1943-1981.* ERA.

[29] Xinyu Chen, Youngwoon Cho, and Suk Young Jang. 2015. Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*. IEEE, 63–68.

[30] Camille Cobb, Ted McCarthy, Annuska Perkins, Ankitha Bharadwaj, Jared Comis, Brian Do, and Kate Starbird. 2014. Designing for the Deluge: Understanding & Supporting the Distributed, Collaborative Work of Crisis Volunteers. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 888–899. https://doi.org/10.1145/2531602.2531712

[31] Clara Crivellaro, Rob Comber, John Bowers, Peter C. Wright, and Patrick Olivier. 2014. 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 (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3573–3582. https://doi.org/10.1145/2556288.2557100

[32] Nick Cumming-bruce. 2017. Ex-Judge Chosen by U.N. to Gather Evidence of Syria War Crimes. https://www.nytimes.com/2017/07/04/world/middleeast/syria-war-crimes-prosecution-united-nations.html?_r=0&module=inline

[33] Dharma Dailey and Kate Starbird. 2017. Social Media Seamsters: Stitching Platforms & Audiences into Local Crisis Infrastructure. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1277–1289. https://doi.org/10.1145/2998181.2998290

[34] danah boyd and Kate Crawford. 2012. CRITICAL QUESTIONS FOR BIG DATA. *Information, Communication & Society* 15, 5 (2012), 662–679. https://doi.org/10.1080/1369118X.2012.678878 arXiv:https://doi.org/10.1080/1369118X.2012.678878

[35] Débora de Castro Leal, Max Krüger, Kaoru Misaki, David Randall, and Volker Wulf. 2019. Guerilla Warfare and the Use of New (and Some Old) Technology: Lessons from FARC's Armed Struggle in Colombia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 580, 12 pages. https://doi.org/10.1145/3290605.3300810

[36] Munmun De Choudhury, Andrés Monroy-Hernández, and Gloria Mark. 2014. Narco emotions: affect and desensitization in social media during the mexican drug war. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 3563–3572.

[37] Centro de Colaboración Cívica and MacArthur Foundation. 2018. El Movimiento por nuestros desaparecidos en México y su camino hacia la incidencia legislativa: La siembra colectiva, una apuesta por la esperanza. http://colaboracioncivica.org/esp/wp-content/uploads/2018/08/CCC_SisMNDM.pdf

[38] Comisión Mexicana de Defensa y Promoción de los Derechos Humanos. 2015. 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.* (March 2015). http://www.cmdpdh.org/publicaciones-pdf/cmdpdh-violaciones-graves-a-ddhh-en-la-guerra-contra\-las-drogas-en-mexico.pdf

[39] Jill P. Dimond, Michaelanne Dye, Daphne Larose, and Amy S. Bruckman. 2013. Hollaback! The Role of Storytelling Online in a Social Movement Organization. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 477–490. https://doi.org/10.1145/2441776.2441831

[40] Matthew Edwards, Awais Rashid, and Paul Rayson. 2015. A systematic survey of online data mining technology intended for law enforcement. *ACM Computing Surveys (CSUR)* 48, 1 (2015), 15.

[41] Sheena L. Erete. 2015. Engaging Around Neighborhood Issues: How Online Communication Affects Offline Behavior. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1590–1601. https://doi.org/10.1145/2675133.2675182

[42] Coral Featherstone. 2013. Identifying vehicle descriptions in microblogging text with the aim of reducing or predicting crime. In *2013 International Conference on Adaptive Science and Technology*. IEEE, 1–8.

[43] Pablo Ferri. 2019. "Después de contar otra vez, la cifra de desaparecidos en México va a aumentar un 30%". https://elpais.com/internacional/2019/12/16/actualidad/1576451333_438814.html

[44] Forbes. 2019. México no tiene una cifra oficial de desaparecidos • Forbes México. https://www.forbes.com.mx/mexico-no-tiene-una-cifra-oficial-de-desaparecidos/

[45] Open Society Foundations. 2016. Atrocidades Innegables: Confrontando Crimenes De Lesa Humanidad en México. *Open Society Foundations* (2016). https://www.justiceinitiative.org/uploads/5d386d17-57aa-4b74-b896-43883af55574/undeniable-atrocities-esp-2nd-edition.pdf

[46] Darwin Franco-Migues. 2015. La opacidad en "el registro" de desaparecidos en México. *Del dicho al hecho: opacidad, autoritarismo y verdades a medias* (Apr 2015). https://doi.org/10.2307/j.ctvd7w90t.12

[47] Óscar Daniel Fuentes Rodríguez. 2017. Historia de la desaparición en México: perfiles, modus y motivaciones. *Derecho y Ciencias Sociales* (Nov 2017). https://doi.org/10.24215/18522971e018

[48] Ryan J. Gallagher, Elizabeth Stowell, Andrea G. Parker, and Brooke Foucault Welles. 2019. Reclaiming Stigmatized Narratives: The Networked Disclosure Landscape of #MeToo. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 96

(Nov. 2019), 30 pages. https://doi.org/10.1145/3359198
[49] Matthew S Gerber. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61 (2014), 115–125.
[50] T Gillespie. 2010. The politics of 'platforms'. *new media & society* 12, 3 (May 2010), 347–364.
[51] Tarleton Gillespie. 2011. Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. *Culture Digitally* (2011).
[52] Christine Grillo. 2017. Locating Hidden Graves in Mexico. https://hrdag.org/2017/03/23/locating-hidden-graves-in-mexico/
[53] abril – septiembre 2018 pp. 162-174. DOI: https://doi.org/10.20318/eunomia.2018.4161 162 FORO Y ÁGORA José Antonio Guevara Bermúdez, . ISSN 2253-6655 Nº. 14 and Lucía Guadalupe Chávez Vargas. 2018. Impunity in the context of enforced dissappearance in Mexico. http://www.cmdpdh.org/publicaciones-pdf/cmdpdh-la-impunidad-en-el-contexto-de-la-desaparicion-forzada.pdf
[54] John R Hipp, Christopher Bates, Moshe Lichman, and Padhraic Smyth. 2019. Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Quarterly* 36, 4 (2019), 718–748.
[55] Juan Pablo Hourcade, Lisa Nathan, Panayiotis Zaphiris, Yoram Chisik, Cuauhtémoc Rivera-Loaiza, and Jason C. Yip. 2016. Conflict & HCI: Preventing, De-Escalating and Recovering. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1127–1130. https://doi.org/10.1145/2851581.2886432
[56] INEGI. 2018. Encuesta Nacional de Seguridad Pública Urbana (ENSU) Tercer Trimestre 2019. https://www.inegi.org.mx/contenidos/programas/ensu/doc/ensu2018_marzo_presentacion_ejecutiva.pdf
[57] INEGI. 2019. Encuesta Nacional de Seguridad Pública Urbana (ENSU) Tercer Trimestre 2019. https://www.inegi.org.mx/contenidos/programas/ensu/doc/ensu2019_septiembre_presentacion_ejecutiva.pdf
[58] Amnesty International. 2011. Deadly detention:Deaths in custody amid popular protest in Syria. https://www.amnestyusa.org/wp-content/uploads/2017/04/mde240352011en_5.pdf
[59] Amnesty International. 2013. How is Social Media Transforming Human Rights Monitoring? https://tinyurl.com/y4wbee2o
[60] Cristina Kadar, Yiea-Funk Te, Raquel Rosés Brüngger, and Irena Pletikosa Cvijikj. 2016. Digital Neighborhood Watch: To Share or Not to Share?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 2148–2155. https://doi.org/10.1145/2851581.2892400
[61] Panos Kostakos. 2018. Public Perceptions on Organised Crime, Mafia, and Terrorism: A Big Data Analysis based on Twitter and Google Trends. *International Journal of Cyber Criminology* (2018).
[62] Sheena Lewis and Dan A. Lewis. 2012. Examining Technology That Supports Community Policing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1371–1380. https://doi.org/10.1145/2207676.2208595
[63] Yanni Alexander Loukissas. 2019. *All data are local thinking critically in a data-driven society.* The MIT Press.
[64] Ixtlixóchitl López. 2019. Familiares exigen participar en la creación de la Ley de Desaparición Forzada en la CDMX. https://www.proceso.com.mx/602679/familiares-exigen-participar-en-la-creacion-de-la-ley-de-desaparicion-forzada-en-la-cdmx
[65] Mirca Madianou. 2019. Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises. *Social Media + Society* 5, 3 (2019), 2056305119863146. https://doi.org/10.1177/2056305119863146
[66] Nick Malleson and Martin A Andresen. 2015. Spatio-temporal crime hotspots and the ambient population. *Crime science* 4, 1 (2015), 10.
[67] Vukosi Marivate and Pelonomi Moiloa. 2016. Catching crime: Detection of public safety incidents using social media. In *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*. IEEE, 1–5.
[68] Kelly McEvers. 2017. YouTube Inadvertently Erases Syrian War Videos In Purge Of Extremist Propaganda. https://www.npr.org/2017/09/13/550757777/youtube-inadvertently-erases-syrian-war-videos-in-purge\-of-extremist-propaganda
[69] Ella McPherson. 2015. ICTs and Human Rights Practice: A report prepared for the UN Special Rapporteur on extrajudicial, summary and arbitrary executions. *Centre of Governance and Human Rights, University of Cambridge* (2015).
[70] Mónica Meltis, Jorge Ruiz, and Patrick Ball. 2017. New results for the identification of municipalities with clandestine graves in Mexico. https://hrdag.org/2017/11/23/new-clandestine-graves-mexico/
[71] Carolina Gómez Mena. 2019. ONU: tragedia humanitaria en México por desapariciones. https://www.jornada.com.mx/ultimas/politica/2019/08/24/onu-tragedia-humanitaria-en-mexico-por-desapariciones-6291.html
[72] Andrés Monroy-Hernández, Emre Kiciman, Munmun De Choudhury, Scott Counts, et al. 2013. The new war correspondents: The rise of civic media curation in urban warfare. In *Proceedings of the 2013 conference on Computer*

*supported cooperative work*. ACM, 1443–1452.

[73] Juan Montes and José de Córdoba. 2018. 'It's a Crisis of Civilization in Mexico.' 250,000 Dead. 37,400 Missing. https://www.wsj.com/articles/its-a-crisis-of-civilization-in-mexico-250-000-dead-37-400-missing-1542213374

[74] Universidad Iberoamericana Ciudad De México and A.C. Comisión Mexicana de Defensa y Promoción de Derechos Humanos. 2018. Violencia y Terror: Hallazgos sobre fosas clandestinas en México. http://www.cmdpdh.org/publicaciones-pdf/violencia-y-terror-hallazgos-sobre-fosas-clandestinas-en-mexico.pdf

[75] United Nations. 2019. Expertos de la ONU deploran que los crímenes de la guerra sucia en México sigan impunes | Noticias ONU. https://news.un.org/es/story/2019/11/1466051

[76] Harold G Nelson and Erik Stolterman. 2012. The Design Way: Intentional Change in an Unpredictable World. (2012).

[77] Diane F. Orentlicher. 1990. Bearing Witness: The Art and Science of Human Rights Fact-Finding.

[78] Sam Gregory Yvette Alberdingk Thijm Padania, Sameer and Nunez Bryan. 2011. Cameras Everywhere: Current Challenges and Opportunities at the Intersection of Human Rights, Video and Technology.

[79] Leysia Palen and Kenneth M. Anderson. 2016. Crisis informatics—New data for extraordinary times. *Science* 353, 6296 (2016), 224–225. https://doi.org/10.1126/science.aag2579 arXiv:https://science.sciencemag.org/content/353/6296/224.full.pdf

[80] Fuerzas Unidas por Nuestros Desaparecidos en Nuevo León. 2019. Un sentido de vida: la experiencia de búsqueda de Fuerzas Unidas por Nuestros Desaparecidos en Nuevo León 2012–2019. https://ibero.mx/files/2019/fundenl-informe.pdf

[81] Megan Price. 2018. When Data Doesn't Tell the Whole Story. https://www.ijmonitor.org/2015/05/when-data-doesnt-tell-the-whole-story/

[82] Claudia E. G. Rangel Lozano and Evangelina Sánchez Serrano. 2015. *MÉXICO EN LOS SETENTA: Guerra sucia o terrorismo de Estado? Hacia una política de la memoria.* Itaca / UAG.

[83] Christian Reuter and Marc-André Kaufhold. 2018. Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics. *Journal of Contingencies and Crisis Management* 26, 1 (2018), 41–57. https://doi.org/10.1111/1468-5973.12196

[84] Shakila Khan Rumi, Ke Deng, and Flora Dilys Salim. 2018. Crime event prediction with dynamic features. *EPJ Data Science* 7, 1 (2018), 43.

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

[86] Mario Santiago Juárez, Juan Carlos Gutiérrez Contreras, and Zúe Valenzuela Contreras. 2017. Guía práctica sobre la aplicación del protocolo homologado para la búsqueda de personas desaparecidas. https://movndmx.org/guia-practica-sobre-la-aplicacion-del-protocolo-para-la-busqueda-de-personas-desaparecidas/

[87] Christine Satchell and Marcus Foth. 2011. Welcome to the Jungle: HCI After Dark. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 753–762. https://doi.org/10.1145/1979742.1979630

[88] Saiph Savage and Andrés Monroy-Hernández. 2015. Participatory Militias: An Analysis of an Armed Movement's Online Audience. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 724–733.

[89] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. 2016. Botivist: Calling Volunteers to Action Using Online Bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 813–822. https://doi.org/10.1145/2818048.2819985

[90] Ben Schneiderman, Marc A.. Smith, N Milic-Frayling, E Mendes Rodrigues, J Leskovec, C Dunne, and A. Cenni. 2010. NodeXL: a free and open network overview, discovery and exploration add-in for Excel. https://www.smrfoundation.org

[91] Alexandra Segerberg and W. Lance Bennett. 2013. *The Logic of Connective Action.* Cambridge University Press.

[92] Irina Shklovski and Volker Wulf. 2018. The Use of Private Mobile Phones at War: Accounts From the Donbas Conflict. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 386, 13 pages. https://doi.org/10.1145/3173574.3173960

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

[94] The Mexican Commission of Defense The Federation for Human Rights (FIDH), Human Rights Promotion (CMDPDH), and The Citizens' Commission of Human Rights of the Northeast (CCDH). 2014. Mexico: Report on the alleged commission of crimes against humanity in Baja California between 2006 and 2012. (October 2014). http://www.cmdpdh.org/publicaciones-pdf/cmdpdh_mexico_report_on_the_alleged_commission_of_crimes_against_humanity_in_baja_california_2006_2012.pdf

[95] George E Tita and Adam Boessen. 2011. Social networks and the ecology of crime: using social network data to understand the spatial distribution of crime. *The SAGE handbook of criminological research methods* (2011), 128.

[96] Zeynep Tufekci. 2017. *Twitter and Tear Gas: The Power and Fragility of Networked Protest.* YALE University Press.

[97] Doria del Mar Vélez Salas and Manuel Alejandro Vélez Salas. 2017. Desapariciones forzadas e involuntarias: El registro estadístico de la desaparición: delito o circunstancia? (Feb 2017). http://onc.org.mx/wp-content/uploads/2017/02/fasciculo-desapariciones_digital.pdf

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

[99] Paulina Villegas. 2018. 168 Skulls in Mexican Mass Grave Are Reminder of Cartel Violence. https://www.nytimes.com/2018/09/06/world/americas/mexico-mass-grave.html

[100] Lizzie Wade. 2017. Mapping Mexico's hidden graves. https://www.sciencemag.org/news/2017/06/mapping-mexico-s-hidden-graves

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

[102] Human Rights Watch. 2011. Neither Rights Nor Security Killings, Torture, and Disappearances in Mexico's War on Drugs. *Human Rights Watch* (November 2011). https://www.hrw.org/sites/default/files/reports/mexico1111webwcover_0.pdf

[103] Human Rights Watch. 2013. Los Desaparecidos de México: El persistente costo de una crisis ignorada. https://www.hrw.org/sites/default/files/reports/mexico0213spwebwcover_0.pdf

[104] Human Rights Watch. 2019. Mexico: Lessons from a Human Rights Catastrophe. https://www.hrw.org/blog-feed/mexico-lessons-human-rights-catastrophe

[105] Human Rights Watch. 2019. World Report 2019: Rights Trends in Mexico. https://www.hrw.org/world-report/2019/country-chapters/mexico

[106] Daniel Wilkinson. 2018. Mexico: Forced Disappearance, an Ongoing Crime. https://www.hrw.org/news/2018/11/26/mexico-forced-disappearance-ongoing-crime

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

[108] Wola. 2018. Mexico Must Summon Political Will to Fully Implement Law on Forced Disappearances. https://www.wola.org/2018/08/mexico-disappearances-crisis/

[109] Donghyeon Won, Zachary C. Steinert-Threlkeld, and Jungseock Joo. 2017. Protest Activity Detection and Perceived Violence Estimation from Social Media Images. In *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*. Association for Computing Machinery, New York, NY, USA, 786–794. https://doi.org/10.1145/3123266.3123282