

---

# “You Are What You Tweet!” The Ethics of (Re) Publishing Public Data as Crafted Narratives

**Melissa Bica**

University of Colorado Boulder  
Boulder, CO 80303, USA  
melissa.bica@colorado.edu

**Jennings Anderson**

University of Colorado Boulder  
Boulder, CO 80303, USA  
jennings.anderson@colorado.edu

**Abstract**

Social media research in HCI often requires deep qualitative investigation, looking beyond the face value of the data to contextualize the story and recognize the people involved. This can muddle the line between behaviors that are considered public, versus those behaviors that need human subjects governance. Here we present a case study of republishing public social media data collected during Hurricane Sandy in the US. We discuss ethical implications that arose from our specific treatment and reuse of the data and the consequent view of participants' tweets as more than public data, which put us at legal odds with terms of service agreements that differently protect people as human subjects and people as content authors.

**Author Keywords**

Data Privacy; Social Media Research Ethics; Twitter

**ACM Classification Keywords**

K.7.4 Professional Ethics: Codes of ethics: Codes of good practice: Ethical dilemmas.  
H.5.m. Information interfaces and presentation: Miscellaneous.

**Introduction**

When content is publicly published on Twitter, what are the ethics surrounding how that data should be interpreted and reused by researchers? Researchers can be dismissive of this issue because the data are publicly visible. We, however, argue that in-depth scrutiny by researchers is likely not a response that authors expect to elicit from their spur-of-the-moment tweets. Conducting in-depth research through a close reading and interpretation of these data requires considering evolving ethical norms in this space.

As *crisis informatics* researchers, we look to understand the collaboration and cooperation of people via social computing platforms in times of crisis and disaster. Because Twitter is a primary social computing platform in this domain, we often study tweets from people experiencing disasters [6]. boyd and Crawford claim that social media “data [are] created in highly context-sensitive spaces, and it is entirely possible that some users would not give permission for their data to be used elsewhere” [2]. As researchers, we must try our hardest to interpret tweets within these contexts. As we will discuss in this piece, reading tweets closely and interpreting them as personal accounts of precautionary actions during a disaster engages complexities in human subjects research. We have experienced and will show that Institutional Review Boards (IRBs) have had to wrestle with the difficult problems around reasonable expectations of publicity and privacy: for now, in the US, many IRBs have determined the data to be public.

#### *Motivation & Background*

During a research methods course in which the authors were involved, the professor organized a discussion around human subjects research for which she brought

in an expert. The class asked the speaker, “What about Twitter?” The speaker quickly responded, “It’s all public, we don’t care.” Imagine the astonishment of the students who were newly learning about research methods for the first time, but who had a lot of their own personal experience with social media.

As researchers of this ethical issue, we know that the “we don’t care” speaks to the issue of *determinations* and of what is *in scope*. Naturally, a human subject expert “cares.” Furthermore, we also know that the quickness of the response masked the decade of inquiry that human subjects experts have conducted on this very issue, and that, for now, this determination of social media and publicity represents just one point in time within a long string of decisions. Nevertheless, this initial discussion had a massive impact on the students in this class.

Just a few weeks later, we brought up our case study (described next). The sociologists in the room were opposed to republishing tweets without taking considerable efforts to anonymize and protect authors. When the class further learned that modifying or anonymizing the tweets would be in direct violation of the Twitter Terms of Service, the class reacted with further disbelief. It was at this point the professor asked, “I wonder what the IRB would say?” Recalling the earlier discussion about publicly available social media data, the class realized that the legalities governing human subjects and the legalities governing social media authorship are in conflict. Furthermore, the representational qualities of social media data are changing, and the analytics HCI researchers can perform are rapidly advancing; thus, today’s human subjects determinations—no matter how well thought

through—cannot be presumed to anticipate even near-future research cases.

### **Our Research & Case Study**

Hurricane Sandy was the second costliest hurricane in US history, causing an estimated \$6 billion in damages after making landfall on the eastern seaboard in October 2012 [1]. It became the most tweeted about event to that date. However, when examining geotagged tweets from the event, we found some of the most affected areas to be under-represented in the social media conversation around the storm. In a recent study, we examined the hurricane-related experiences and behaviors shared on Twitter by residents of Far Rockaway, a coastal New York City neighborhood that is geographically and socioeconomically vulnerable to disasters and was significantly affected by the storm.

We collected the entire tweet histories for users mentioning “Far Rockaway” or “Farrock” in their tweets or profile descriptions during the storm. After manually filtering these tweets to include only those from users who we determined were residents of Far Rockaway, we qualitatively coded about 7K tweets by 41 users for personal sentiment, preventive actions, and information seeking and sharing practices. We found the tweets to offer rich personal accounts of Twitterers’ experiences with Sandy. Analyzing these first-person narratives, we identified risk perception and protective decision-making behavior. We also found themes of invisibility and neglect as residents expressed feeling abandoned by the media, the city government, and the overall relief efforts in the aftermath of Sandy. At the end, we created “thick descriptions” [3] for two users who exemplify major themes of both protective decision-

making and frustration with local media and government over a lack of representation.

This research was recently accepted to a conference on information systems for crisis response [1], which has motivated us to reflect on how we should portray the two personal narratives we have constructed.

### **Our Legal and Ethical Quandary**

In creating our “thick descriptions” for the two exemplary users in our collection, as students doing research of this kind for the first time, we came to understand that we were more than simply republishing their public social media posts—we have closely interpreted their posts in the context of their full post histories and profiles to craft narratives about their experiences during Hurricane Sandy. The reasoning “but the data [are] already public” [7] did not feel like enough to legitimize our retelling of their story from the data.

#### *The Legal Complications*

Since we have collected all of our data for this study through Twitter, we had a responsibility to adhere to their terms of service. Therefore, we first turned to Twitter to gain insight on our ethical dilemma from a legal standpoint. Anyone who uses Twitter services is required to agree to both their Terms of Service and Privacy Policy. A callout at the top of both of these documents reads: “*Tip: What you say on the Twitter Services may be viewed all around the world instantly. You are what you Tweet!*”<sup>1</sup> Here, Twitter makes it clear that once a public tweet is posted, authors should not expect privacy or protection (other than authorship)

---

<sup>1</sup> Twitter Terms of Service. <https://twitter.com/tos>

from the company.

Furthermore, Twitter's Privacy Policy continues to even list potential uses of the data:

*The Twitter Services broadly and instantly disseminate your public information to a wide range of users, customers, and services... public Tweets are immediately delivered... to our partners... including search engines, developers... and institutions such as universities and public health agencies that analyze the information for trends and insights. When you share information or content like photos, videos, and links via the Services, you should think carefully about what you are making public.<sup>2</sup>*

Additionally, Twitter provides guidelines to people who wish to use their data. Their Display Requirements<sup>3</sup> instruct anyone who displays tweets to "Display real, unmodified Tweets from real accounts." Following these terms, then, we are obligated to display the full tweet content including full name and Twitter username.

Twitter not only provides ample warning to all users about their privacy rights and the potential uses of their data, but instructs researchers who collect and use this data to do so in a personally-identifiable way. While the official Twitter privacy stance is clear, we feel compelled as human computer interaction researchers to consider whether users' acceptance of these click-through [4] policies and guidelines denotes their informed consent of our critical examination and republishing of their data. Thus, we consider not only

---

<sup>2</sup> Twitter Privacy Policy. <https://twitter.com/privacy>

<sup>3</sup> Twitter Display Requirements. <https://about.twitter.com/company/display-requirements>

the legality of using Twitter data, but the ethics. Just as we have accountability toward Twitter itself, we also have accountability toward our participants—the users whose tweets we include in our study. We must consider people's rights—not only as Twitter users, but as people who publish content in general—and what they view as "acceptable reuse of their data" [2]. Here we find a disconnect between the rights of social media users and the legal guidelines provided by Twitter [8]. Ideally, there would be an overarching standard for researchers to adhere to in this situation, but even IRBs, the reigning authority on ethical human subjects research, do not agree on this matter, or even have an opinion at all in some cases [5].

### **Six Ethical Questions of Our Research**

In light of these legal guidelines, we came up with six questions regarding our study that we felt needed to be addressed from an ethical standpoint in addition to, or perhaps in place of, a legal one.

*(1) Should we ask for consent from all users whose tweets we plan to publish in our research?*

In general, we feel that since the posts we used were publicly available when originally posted and still are at the time of publication, we do not need to ask for consent from users to use their tweets. However, given the in-depth nature of our study, we believe the response to this question also depends heavily on the amount of data that we choose to publish for a given user as well as whether we are using the tweet text at face value, or performing a deeper interpretation to infer and expose more implicit meaning.

For most of the users in our study, we carefully selected individual tweets that are exemplary of larger

themes, such as preparation or information seeking. For example, we used the following tweets as examples of these two categories, respectively:

***infinetwords14*** (Oct 28 11:57): *House is stocked up, got the 1st aid kits, bottled water. Now, we wait. Hunkered down for Sandy.*

***CHDADONNRF*** (Oct 31 21:13):  
*@DanielSquadron @NYCHA @chriscquinn @rosiemendez #sandy what's going on with NYCHA in far rockaway. Are they safe from looting?*

Interpreting just the text of these posts in the context of Sandy is straightforward—the first tweet clearly presents a description of one user’s preparedness for the impending storm, while the second asks for information regarding safety in the hurricane’s aftermath. Therefore, we chose to use these and other similar posts in our study without requesting consent from the users.

*(2) Should we contact the two users whose posts we used to create the thick descriptions in our publication?* We did choose to contact the two users whose tweets we used to compose descriptive narratives. Although we still carefully selected these tweets and avoided using sensitive content, we included a more substantial amount of data from these users in our study. Rather than using one or two tweets to represent a single theme or idea, crafting detailed narratives required us to interpret these tweets according to the context of the users’ other tweets and profiles. The degree to which we investigated this data and built the user stories is likely beyond the original authors’ expectations. We are no longer using the public data at face value, but creating something greater in

composing these stories, which arguably requires different ethical approaches for reuse.

*(3) How can we best approach these users about republishing their tweets?*

We sent these two users carefully crafted messages using the language of informed consent and asking for permission to use and republish their personally identifying data, as well as explaining the purpose of the research and why the inclusion of their data is particularly important to our analysis. We also included a copy of the paper for them to read in full. The message outlined potential risks and benefits of allowing us to use their unmodified tweets, and options for anonymization if they did not want some or any of their data attributed to them. Both users consented to full use of their data.

*(4) What should we do if a user does not want their tweets attributed to them?*

Although we received full consent from users in this particular study, this is still an issue for future studies of similar nature. If users do not consent, this puts us in the position to choose between their preferences and the social media platform’s established legal requirements. In our outreach message, we explicitly stated that the user had anonymization options. Although anonymizing tweets could disobey the Terms of Service, we felt that in this case the ethics outweigh the legalities—it was more important to allow the users to choose how to be represented in our research accurately and appropriately. Since neither user we contacted chose to anonymize their data, this remains an open ethical question: should users be allowed to choose whether to have their data anonymized in research studies, and if so, to what extent?

(5) *What should we do if the users do not want their tweets republished at all? Should this be an option?* Similarly, we did not need to take action on this issue, so it, too, remains open. The ultimate question here is: should people be allowed to refuse others the right to use their previously published data? Should this ever become an issue, it will be difficult to defend because the public nature of the data is made very clear by the platform.

(6) *What are the risks and/or benefits of publishing identifiable information, such as usernames, in our research?*

Including identifiable information has no direct impact on the contribution of our research. However, we are concerned about this issue because we see potential risks and benefits for the participants. Specifically, risks could include potential exposure of personal, identifiable information, and possibly sensitive content, though we have tried to be responsible to exclude tweets that expose personal information. On the other hand, we see this research as an opportunity for participants to share their stories and experiences during Hurricane Sandy with a wider audience, something that was a challenge for Far Rockaway residents when the storm occurred.

### **Conclusion**

The ethical practices surrounding the reuse of public social media data are uncertain with different stakeholders having varying views; it is unclear which view trumps the next. Because HCI researchers have the techniques to investigate data deeply and broadly, the ethical issues are renewed. In the end we find that the legalities around people *as human subjects* versus people *as authors* come into conflict, with both views in

contention around what best protects people and what is viewed as people's intellectual property. It seems that the issue must currently be debated on a case-by-case basis.

### **Acknowledgements**

We thank our colleagues for their support. We also thank our study population, most of whom are not aware of our research efforts.

### **References**

1. Jennings Anderson, Marina Kogan, Melissa Bica, Leysia Palen, Ken Anderson, Kevin Stowe, Rebecca Morss, Julie Demuth, Heather Lazrus, Olga Wilhelmi, and Jen Henderson. 2016. Far Far Away in Far Rockaway: Responses to Risks and Impacts during Hurricane Sandy through First-Person Social Media Narratives. To appear in *Proc. of ISCRAM 2016*.
2. danah boyd and Kate Crawford. 2012. CRITICAL QUESTIONS FOR BIG DATA. *Information, Communication & Society*, 15(5), 662-679. doi:10.1080/1369118X.2012.678878
3. Clifford Geertz. 1973. Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures: Selected Essays*. Basic Books, New York, 3-30.
4. Nathaniel Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz, Deirdre Mulligan, and Joseph Konstan. 2005. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *Proc of the 2005 symposium on Usable privacy and security (SOUPS '05)*. ACM, New York, NY, USA, 43-52.
5. Heather Small, Kristine Kasianovitz, Ronald Blanford, and Ina Celaya. 2012. What Your Tweets Tell Us About You: Identity, Ownership and Privacy of Twitter Data. *International Journal of Digital*

*Curation*. 7, 1: 174-197.

6. Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging During Two Natural Hazards Events : What Twitter May Contribute to Situational Awareness. In *Proc of CHI 2010*. <http://doi.org/10.1145/1753326.1753486>
7. Michael Zimmer. 2010. "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology*. 12, 4: 313-325.
8. Michael Zimmer and Nicholas Proferes. 2014. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Info Mgmt*. 66, 3: 250-261.