Being "Public" while Vulnerable: Recommendations for Researchers and Platforms

Casey Fiesler
University of Colorado Boulder
Boulder, CO
casey.fiesler@colorado.edu

Abstract
Though both research ethics heuristics and privacy design are often one-size-fits-all, vulnerable communities interacting online have specific, often non-obvious concerns when it comes to both. Ongoing work in my lab highlights these concerns and points to some recommendations for both research methods best practices and platform design and policy.

Author Keywords
ethics; platforms; privacy; research ethics; vulnerable communities

CHI 2020 Workshop on Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations

Introduction
Data collection and use is a growing concern with respect to privacy, even in the context of scientific research. However, currently common research ethics heuristics are one-size-fits-all when it comes to privacy concerns; whether the data "public" or not [12,13], regardless of who created it and whether they might be in a vulnerable position.

Ongoing research in my lab speaks to research ethics generally when it comes to this type of public data collection, as well as to the specific concerns relevant to vulnerable, marginalized, or stigmatized communities, such as LGBTQ people, racial minorities, or health support communities. Collectively, this work suggests reasons why both privacy design and research practices should *not* be one-size-fits-all, and also points to some specific recommendations for research methods and platform design and policy.

Recommendations for Researchers

The theory of contextual integrity provides that an appropriate flow of information for the purposes of privacy will conform to contextual norms [9]. These norms might change according to characteristics such as content, parties involved, and evolving ethical. In a study of how Twitter users feel about researchers collecting and analyzing tweets, we saw evidence of how this kind of context matters when it comes to comfort levels with research—for example, data analysis methods, how data will be reported, and of course, the nature of the data itself [6]. In simple terms, a tweet about what someone had for breakfast is not equivalent to a tweet revealing someone's HIV status, for the purposes of how someone might feel about its collection and use. Though this might seem obvious, if our only ethical guideline is "is the data public," there is nothing to suggest that, for example, the latter tweet should not be quoted in a paper. Research has shown that even for specifically health-related articles, a large number directly quote tweets and of those, the vast majority lead back to the user who posted it [1].

In addition to other issues (e.g., I didn't give consent; I don't want my content taken out of context, or even objecting entirely to the notion of being researched [7]), many of the concerns expressed in this study were directly privacy related—such as wanting an assurance of anonymity. These concerns are particularly important for people who are in vulnerable positions and might, for example, be online and "public" because they are seeking social support. It is therefore critical that data collection, analysis, and reporting decisions take into account both the nature of the content itself and the circumstances of the person who created it.

In a study of how people in a majority LGBTQ community feel about uses of public data by both researchers and journalists, we heard about complex privacy concerns that might not be obvious from the content alone [4,5]. For example, if artwork from someone's Tumblr were shared in a news article, what if someone recognized their art style and this led to their pseudononymous Tumblr where they were "out" in a way they were not in their physical life? Many participants expressed concern about the amplification of their content beyond the context in which they originally shared it.

Taken together, these concerns emphasize the importance of researchers considering these contextual factors and making appropriate methodological choices. For example, when a tweet is about a sensitive topic, it should not be quoted in a paper (particularly since it could make its way to the media [6]). In those cases, omission or ethical fabrication [8] would be appropriate out of an abundance of caution for the content creator's privacy. But the important recommendation overall here is that researchers must take into account context beyond "publicness" of data in making decisions about how to collect, analyze, and report that data—and one of those pieces of context is the vulnerability of the people who created it.

Recommendations for Platforms

A common response to these kinds of concerns is that because the data is "public," people should be aware that it can be used by anyone for any purpose. However, we know that there is a knowledge gap about

the ways that information can flow *beyond* a platform like Twitter [11], that people rarely read TOS and privacy policies [10], and (as revealed by our study of Twitter users) that most people likely have no idea that researchers make use of public content like tweets [6].

Moreover, some people object entirely to their content being used beyond the context of the platform at all without their consent [7]. However, for the most part, if you don't want your content on a platform to be used in a certain way, your option is to not use that platform. Similarly, on platforms without granular privacy settings—e.g., Twitter and Tumblr—your only option is to be *all* public or *all* private. This choice is a difficult one for communities that are both vulnerable to attack *and* where social support is critically important.

Our same participants who worried about the amplification of their content also worry about limiting the discoverability of their content. While privacy is important, it is also the case that when people may not have a support system in their physical space, being able to "stumble" onto a supportive online community (e.g., without having to risk googling "queer support group") can be life-changing [3]. This creates a *very* complicated design space for privacy—and also a non-obvious one. How do you simultaneously design for both privacy and discoverability?

Our research suggests a few design possibilities, but even these would not make sense under all circumstances with all groups. For example, granular privacy settings (where visibility can vary from post to post) allow for decisions about appropriate information flow based on a single piece of content rather than an entire account. Individuals' online lives, even a single platform, are not limited to a single context. One Tumblr post might be a piece of art that they want to publicize, and the next might be a personal story about gender transition or disability. Designing privacy settings so that *some* content is "public" but other content is not reduces anxiety about content randomly showing up on Buzzfeed.

Relatedly, different subcommunities on a single platform might have different content-based contexts and different norms about privacy. Tumblr and Twitter flatten communities so that everyone interacts in the same space whether they want to or not. By contrast, LiveJournal includes "communities" that are similar in structure to Usenet groups or subreddits but that allow group-based privacy settings so that the content can be invisible to outsiders. A similar option is to make content visible only to logged in users; though this too can go wrong (e.g., when someone spoofed a logged-in user to scrape all of OKCupid [13]) because definitions of what constitutes "public" vary [2]. However, sub-communities also allow for social norms to create and enforce privacy practices—and also for community members to help each other be aware of policies, design affordances, and presence of outsiders.

Conclusions and Ongoing Work

"Context matters" is a simple recommendation when it comes to privacy, but critically important in the context of vulnerable communities—and support for context can be built into both best practices and design. Next steps in this work include consideration of a platform where design can be controlled for while considering ethics and privacy attitudes of different types of communities (Reddit), and a community where researcher positionality is a particularly relevant context (Black

Twitter). As highlighted in our prior work with an LGBTQ community [5], the best path towards both researching and designing for a vulnerable community is to understand their context, concerns, and needs.

References

1. John W. Ayers, Theodore L. Caputi, Camille Nebeker, and Mark Dredze. 2018. Don't quote me: reverse identification of research participants in social media studies. *npj Digital Medicine* 1, 1: 29–30.
2. J.C.H. Bromseth. 2002. Public places - public activities? Methodological approaches and ethical dilemmas in research on computer-mediated communication contexts. In *Resesarching ICTs in Context*, A. Morrison (ed.). University of Oslo, 33–61.
3. Brianna Dym, Jed R. Brubaker, Casey Fiesler, and Bryan Semaan. 2019. Coming Out Okay: Community Narratives for LGBTQ Identity Recovery Work. *Proc. ACM Human-Computer Interaction* CSCW.
4. Brianna Dym and Casey Fiesler. 2018. Vulnerable and Online: Fandom's Case for Stronger Privacy Norms and Tools. In *Companion of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, 329–332.
5. Brianna Dym and Casey Fiesler. 2020. "First rule of fandom": Ethical and privacy considerations for research using online fandom data. *Transformative Works and Cultures*, forthcoming.
6. Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1.
7. Blake Hallinan, Jed R Brubaker, and Casey Fiesler. 2019. Unexpected expectations: Public reaction to the Facebook emotional contagion study. *New Media & Society*: 146144481987694.
8. Annette Markham. 2012. Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society* 5, 3: 334–353.
9. Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79: 101–139.
10. JonathanA. Obar and Anne Oeldorf-Hirsch. 2018. The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. *Information Communication and Society*: 1–20.
11. Nicholas Proferes. 2017. Information Flow Solipsism in an Exploratory Study of Beliefs About Twitter. *Social Media + Society* 3, 1: 205630511769849.
12. Michael Zimmer. 2010. "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology* 12, 4: 313–325.
13. Michael Zimmer. 2016. OKCupid Study Reveals the Perils of Big Data Science. *WIRED*. Retrieved from https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/