What Is "Publicly Available Data"? Exploring Blurred Public–Private Boundaries and Ethical Practices Through a Case Study on Instagram

Signe Ravn[1] iD **, Ashley Barnwell**[1]**, and Barbara Barbosa Neves**[1]

Abstract

This article adds to the literature on ethics in digital research by problematizing simple understandings of what constitutes "publicly available data," thereby complicating common "consent waiver" approaches. Based on our recent study of representations of family life on Instagram, a platform with a distinct visual premise, we discuss the ethical challenges we encountered and our practices for moving forward. We ground this in Lauren Berlant's concept of "intimate publics" to conceptualize the different understandings of "publics" that appear to be at play. We make the case for a more reflexive approach to social media research ethics that builds on the socio-techno-ethical affordances of the platform to address difficult questions about how to determine social media users' diverse, and sometimes contradictory, understandings of what is "public."

Keywords

digital ethics, Instagram, public/private, consent, socio-techno-ethical affordances, research ethics

Introduction

Despite the emerging debate on the ethics of social media research, scholars still tend to rely on technical notions of "publicly available data" when evaluating and developing ethical standards for their research on such platforms (Markham & Buchanan, 2012; Markham, Tiidenberg, & Herman, 2018). Based on the assumption that social media users have knowingly published the content in the public domain, this "consent waiver" approach paves the way for wide-ranging harvesting and analysis of online content as empirical data. In this article, we problematize these assumptions and make the case for a more reflexive approach to social media research ethics. To do this, we explore ethical issues and tensions in our research project "Captured and Captioned: Representing family lives on Instagram" and illustrate how we developed a set of practices to ensure that this research was ethically responsible. We suggest a flexible and responsive form of ethical research that builds on the social and technical affordances of platforms—which we name socio-techno-ethical affordances—to address difficult questions about how to determine social media users' diverse, and sometimes contradictory, understandings of what is "public."

While social media platforms differ considerably with regard to their socio-technical affordances, this article focuses on Instagram, a platform with a distinct visual premise that involves a unique set of concerns. Reproducing users' often quite personal photographs in publications and public presentations, for example, poses particular ethical challenges. We offer our specific case study to explore ethical tensions when studying social media in a way that has relevance for other platforms. We illustrate how ethics is not just a procedural element of ethics applications. In line with Markham and colleagues (Markham et al., 2018), ethics is embedded in the research methods; it must be an ongoing concern during data collection, analysis, and dissemination. In this article, we argue for a more nuanced understanding of how users negotiate the blurred boundaries between public and private in online spaces, with the aim to inform best practices for digital research and advance the discussion on this topic.

Research Ethics and Social Media Research

Researchers have long been grappling with how the Internet and its socio-technical affordances question traditional regulatory concepts such as privacy and the boundary between

[1]The University of Melbourne, Victoria, Australia

Corresponding Author:
Signe Ravn, School of Social and Political Sciences, The University of Melbourne, Parkville, Melbourne, Victoria 3010, Australia.
Email: signe.ravn@unimelb.edu.au

public and private (Markham et al., 2018). For instance, boyd (2008) has problematized the binary logic underlying how platforms such as Facebook think of privacy in simple, technical terms—"data are either exposed or not" (boyd, 2008, p. 16). In her view, privacy is instead about users' sense of control over what information is shared with whom and in which context. Conversely, such control is lost when different contexts are collapsed into one (boyd, 2008; Marwick & boyd, 2010). These issues are of course not isolated to Facebook, but are applicable across social media platforms, and may vary across platforms, depending on how the user negotiates notions of privacy on each platform. Social science research on Instagram has identified a number of ethical challenges related to collecting, analyzing, and reporting data from this platform specifically (Highfield & Leaver, 2015; Highfield & Leaver, 2016; Locatelli, 2017). Instagram, like a number of other platforms, has settings that allow users to control their audience and thereby actively decide what they make publicly available. These settings have developed over time to allow users more privacy options. For instance, adding the direct message function (2013) allowed users to only share content with particular contacts (Highfield & Leaver, 2015). In 2016 the "Stories" function was added, allowing users to post content that is only visible for 24 hours. Researchers can in principle rely on these settings to engage with "publicly available content" when conducting research—on Instagram this means that only accounts that are set to public will show up in searches. If data are deemed to be public, or "naturally occurring" (Silverman, 2007), seeking informed consent from users is not necessary according to most institutional ethics committees.

However, the definitions of public and private are complex, nuanced, and dynamic: users might be revealing more information than expected through the image's background or through the combination of visual and textual elements. Or they might be sharing content of others that has been acquired illicitly such as "creepshots," that is, nonconsensual sharing of pictures (Highfield & Leaver, 2016). The development of Instagram over time—from iPhone users only to other smartphone users and then on the web—means that early adopters of the platform may have their content "exposed" to more and more users (Highfield & Leaver, 2015). Furthermore, sharing one's posts across different platforms means that different privacy settings apply. As we will return to below, our study suggested that the line between public and private was deliberately blurred by a number of users. This emphasizes that relying on a simple understanding of "publicly available" is not sufficient for social media research to be ethical (see also Locatelli, 2017 and Tiidenberg, 2018b for a discussion of this). Responding to these concerns, researchers studying platforms that combine text and image typically employ two main strategies. The first is to exclude the visual material (i.e., pictures)

completely and either describe it alongside the textual material such as caption, hashtags, and comments (Tiidenberg, 2018a) or to aggregate the textual material into critical themes (Friedman, 2018). The second strategy is to use the visual material but cover identifying features, such as eyes, with a black box/censor bar in an attempt to protect users' identity (Caruso & Roberts, 2017). We welcome these efforts to not simply assume public availability but also believe that they have certain limitations. First of all, our study analyzes how people construct family resemblances, not only using physical likenesses but also matching clothes or haircuts to "practice" family life (Morgan, 2011). The visual materials are a key component of analyzing this topic, so excluding images or covering up parts of the images was not a viable approach for us. Second, some scholars argue that modifying people's images is as unethical as showing the full image (Gross, Katz, & Ruby, 2003). And finally, even when researchers make attempts to protect users' privacy by hiding the data source or de-anonymizing participants, we can never guarantee that users cannot be traced and identified via digital traces (Zimmer, 2010).

There are no specific guidelines for Instagram research, but the Association of Internet Researchers (AoIR) advocates for a "case-based" perspective rather than a prescription of procedures (Markham & Buchanan, 2012). Zimmer outlines respecting contextual understandings of privacy and developing data anonymization strategies as important practices (2010) and Highfield and Leaver (2016) argue that instead of focusing on public–private binaries, researchers should consider the potential to do harm, for instance via the representations of research findings. While this is certainly a key consideration, we agree with Markham and Buchanan that a critical understanding of the contextual meanings and axiological implications of "publicly available data" is still lacking in practice (Markham & Buchanan, 2012; see also Zimmer, 2010). This void is a key motivation for the present article. A critical discussion of the current definitions and uses of the notion of "publicly available data" is vital for researchers working on social media platforms that offer a range of dynamic affordances, and where users have diverse understandings about how their data are being displayed and/or accessed by third parties, such as researchers. In the following section, we briefly introduce sociological theories of "the public" to ground our later discussion in the literature.

Advancing the Notion of "The Public"

In *Camera Lucida* (1980), Roland Barthes locates the blur of public and private at the very advent of photography. He writes that "the age of photography corresponds precisely to the explosion of the private into the public, or rather into the creation of a new social value, which is the publicity of

the private: the private is consumed as such, publicly" (Barthes, 1980, p. 98). This description is prescient for digital platforms like Instagram where intimate, family photos are shared and become part of broader aims to publicize identity, or more literally, brands, and businesses. On Instagram very different understandings of what is public, and what account holders see as the specific public for each post, seem to operate. Marwick and boyd (2010) have suggested the term "imagined audiences" to conceptualize how social media users engage in self-presentations and impression management on social media. Here, we want to take this one step further by turning to Lauren Berlant's work, which is helpful for our focus on intimacy.

Berlant's (2008) notion of the intimate public is useful for capturing the nuances of how people define their publics online and has been popular in recent social research on various digital platforms (for instance, Andreassen, 2017; Kanai, 2017). Rather than opposing the intimate and the public, Berlant (2008) conceives of the "intimate public," often produced by media and marketing, as "constituted by strangers who consume common texts and things" (2008, p. viii). For Berlant, this public can be characterized by emotion and personal life, in ways that unsettle more traditional ideas of "public life" that emphasize collective action and civic and democratic debate (Habermas, 1962; Sennett, 1977). What we take from Berlant's notion of the intimate public is that people often presume they know not only the common feeling but also the bounds of their audience. This is a very important consideration for digital researchers who may read a series of posts as performances of political identity when they were not intended as political, or indeed as data for research. As scholars, we may work with shared definitions of "the public," or with legal or institutional cutoffs for what such notions of "public" will allow, but this does not mean that the spaces and people we are researching share this view or knowledge. It is necessary for us to ask, every time, what is the intended public here? And often, as our own efforts indicated, this is very difficult to discern without directly asking the owner of the account.

Recent work on intimate publics and digital culture also raises important ethical questions around how markets and platforms co-opt, represent, and reframe intimate data that are displayed online. In these contexts, for example, Dobson et al. argue that, ironically, "public intimacies are not public enough in the sense that participants have little control over what platforms do with their intimate relationships" (2018, p. 22). Scholars scraping "intimate data" without permission, or without engaging the producers of such data, might be accountable to similar ethical questions or concerns. With this emerging research on what "public" means on social media platforms, researchers are called to develop more nuanced understandings of the "layers of publicness," which Locatelli describes as "ranging from the unrestricted

publicness of viral images to the closeness of private profiles" (2017, p. 3). Working with often sensitive family photography, our process of developing ethical protocols has been both frustrated and enriched by these tricky questions around how publics are defined and by whom, and how these often-blurred boundaries affect the collection, analysis, and representation of social media data.

The "Representing Families on Instagram" Study

We designed our project in 2016 with plans to use web-scraping software to collect data from the platform. But when the project began in 2017, Instagram had changed its setup (API) from allowing automated public collection of data to a set of restricted options (mostly aimed at the IT industry) as well as modifying its terms of use (limiting web scraping). However, Instagram still provides access to "publicly" available data, including personal information on its users. We therefore changed the project design from a mixed methods project to a project that was qualitative in its approach and manually collected 150 posts (photo + accompanying text) related to family relationships and moments. These posts were found by searching specific hashtags such as #familymemories, #motherdaughter, #fatherson, #fatherdaugther, and #motherson. We used a snowball sampling strategy to broaden our search from these popular hashtags to ensure that diverse family representations were also included, for instance leading us to hashtags such as #modernfamilies and #gayfamilies.

Data (*n* = 150 posts) were collected in August to September 2017. When revisiting the data in 2018, some posts were no longer searchable, because the user had deleted the post, changed the account name, changed the privacy settings, or deleted the profile altogether. For ethical reasons, we decided not to include these posts in our sample and had to supplement and boost the initial sample, from July to September 2018. The final sample includes 200 posts in total, which obviously remain open to the same vulnerabilities. Nonetheless, our guidelines were to exclude posts no longer publicly available at the point of starting the analysis process (coding), and after this stage, we kept all posts in the sample for analysis. We only reproduce posts for which we are granted permission by users, also in the rare case that posts were no longer publicly available online.

The project was approved by the University of Melbourne Human Ethics Advisory Group. A key part of the ethics application was ensuring that only publicly available data would be collected. However, as the research developed and we started thinking about how to analyze and present the data, our initial approach to research ethics seemed insufficient. To give a brief account of the context in which our thinking around ethics developed, we will now outline the challenges we encountered and how we sought to

resolve them. We conceptualize the challenges we faced as "ethically important moments" (Guillemin & Gillam, 2004). These are moments that capture unanticipated and challenging situations that emerge in the daily practice of doing social research, but are often overlooked or simply not reported. The "ethically important moments" we experienced in our project evolved from two main aspects, both relating to the question of what is seen as constituting "publicly available" data.

The first of these moments concerns the generally intimate and potentially sensitive nature of family photographs. When considering the posts we had collected for analysis, it quickly became clear that the intimate and sensitive nature of our sample—given the overall focus on intergenerational family moments and memories—posed specific challenges. Most posts included children, and a number of posts included deceased or seriously ill people. Reproducing these posts in academic presentations and publications seemed unethical. These reflections formed an "ethically important moment" in that they made us question reproducing Instagram posts more generally. Despite their technically "public" nature, the posts appeared to publicize "private" content, aimed at friends and/or family members, or specific intimate publics.

The second ethical moment relates to account settings. As noted, when reengaging with the data in 2018, we found some users had set their accounts to private after we contacted them about the project[1]. This made us wonder how many of the users in our sample were actually aware that their privacy settings allowed not only family and friends to see their posts but also "complete strangers" like us, as academics who may be interested in the post for its visual/esthetic qualities and also for analytical reasons. Ultimately, the question that arose for us was whether "publicly available" also means "available for academic scrutiny" (or commercial purposes, for that matter; cf. Highfield & Leaver, 2016). To complicate this even further, when we started to go through the collected posts, we found that some accounts appeared to be professional profiles used to promote various commercial endeavors. These users were mixing family-related posts with posts about their business, thereby posing other questions about account holders' own deliberate blurring of public/private boundaries and if or how this question of "purpose" might affect our ethical assessment.

Moving Forward

To address these issues, we developed several strategies. This process was iterative, with new strategies emerging from both successful and unsuccessful results of previous strategies as we worked through the ethical issues.

First of all, like Locatelli (2017) we did not want to represent users who had not granted permission. As Berlant's notion (2008) of "intimate publics" illuminates, people have different understandings of what they feel as being "their" public. We therefore decided to seek permission from account holders to use the specific post we had collected. We did this by contacting them via the direct message function of the Instagram application. Initially, we only contacted those whose images we expected to reproduce in academic outputs, and had a low response rate. We therefore decided to go through the entire sample of 200 posts and expand our message to include a link to the specific post (so users did not have to scroll back through their archive), and a link to our research blog, as a means to verify our researcher identities. This resulted in the permission to use the post from 50 users. When reproducing these posts in our research outputs, we will anonymise the username to make the post less easily identifiable and searchable.

In the meantime, we also considered alternative ways of representing the Instagram posts in our outputs in case we could not get permissions. We explored ways of creatively representing images in a way that de-identified the people pictured such as "cartoonifying" the images (via a simple online generator). This preserved the general composition of the post but not the detailed likeness of people pictured. While useful to some extent, this solution suffered the same limitations as the "censor bar" approach mentioned earlier, because our study seeks to analyze constructions of "family resemblance."

In addition, we also tried—although sometimes only speculatively—to determine the "purpose" of the account, to consider whether we could use some images without the user's consent. This was particularly the case for accounts that had websites or blogs linked to them as these appeared to (partly) serve commercial or promotional functions for small businesses, such as fitness or yoga teachers. We tried to explore how the account holders framed and designed the publics for their own accounts in different, if subtle, ways. In our efforts to discern what the account holder considered to be their "public," we examined details such as the account name and description, other posts, and any links to external websites. This research was helpful in unpacking some of the complexities around how people determine and frame the audience for their Instagram posts, but ultimately it was too difficult to ascertain the purpose without speaking directly to the user. Even if this had been possible to determine, we would still be left with our general ethical challenge concerning the potentially sensitive nature of the photo, including children being present in the post.

Discussion

Our initial engagement with the data at hand questioned the simple notion of "publicly available data" on Instagram.

Rather than just assuming that these photographs are intended to be public because they are publicly available in terms of being discoverable, we have argued for a more sophisticated understanding of "the public," and following from this, a more considerate approach to using and reproducing data from social media platforms. Drawing on Berlant's work (2008), we can conceive of Instagram as consisting of publics that *feel* shared but are actually diversely defined by different actors. To make ethical decisions, we need to understand what users see and intend as their public and how their posts are addressed to this public. Ascertaining this is not simple. In our opinion, it requires asking the users, as the act of analyzing and reporting findings opens the post up to new "publics" or audiences not intended by the account holder.

While conventional ethical approaches can be limiting for the rapidly evolving field of digital research, particularly due to the nature of different media and the scale of data collection, scholars must still engage with principles of consent, privacy, and ownership. This is also true, we contend, when researching content that is technically publicly available. We are not suggesting an approach that requires "informed consent for everything" (see Salganik, 2018 for examples of cases when informed consent represents risks to participants or is impossible to obtain); rather, we are questioning assumptions that dismiss any form of consent based on notions of data being "publicly available" and thus fair game. This becomes particularly salient in cases like the present one where the data—Instagram posts—to a large extent include children. Locatelli (2017) discusses how the parents in her study had different (or no) strategies for protecting the identity of their child in their Instagram posts. In our study, we are relying on parents consenting to representing their child, and thereby taking responsibility for use of their image.

Our approach means that we are only able to engage with the data at the *aggregated* level (as Highfield & Leaver, 2015, also suggest), but not reproduce the large majority of the posts in our sample. While this can be frustrating, it is no different from conventional ways of working ethically to ensure research participants' anonymity in qualitative research. This sometimes requires omitting revealing details despite their analytical lure. What we are suggesting can be seen as a more sensitive approach to using data from social media platforms, liking it to other types of qualitative data, where protecting our participants is a key principle. As Tiidenberg suggests, "Just because something is technically accessible and collectable, doesn't mean it should be accessed and collected" (Tiidenberg, 2018b, p. 472). Rather than trying to de-identify posts as much as possible, as Highfield and Leaver (2015) suggest, we argue that we must engage with users and ask for consent before reproducing and representing their content as this means altering the audience they are exposed to. This takes seriously their understanding of "intimate publics."

Digital platforms evolve rapidly, partly in response to user needs and concerns. In addition, various social media platforms have very different affordances. When platforms develop, this often entails a change or expansion of the *sociotechnical affordances* of social media—that is, the social and technical cues and means provided by the platforms that shape use and interaction. In line with this, and as shown in this article, we suggest thinking about the socio-techno-*ethical* affordances of the platforms we are working on, that is, drawing actively on the inbuilt functions in apps and platforms such as direct messaging options, filters, and hashtags to underpin our research engagements with the platforms ethically. While this article has focused on Instagram, and a particular set of ethical challenges presented by this platform's visual affordances, other platforms have other socio-techno-ethical affordances that can be identified and activated to ensure ethical engagement. We hope other scholars will take up this task and engage in the discussion about how we can conduct ethically sound and responsive research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Note

1. We can only guess at what prompted this, but this was right after the Cambridge Analytica scandal, which may have made users reconsider the privacy of their social media accounts.

ORCID iD

Signe Ravn 🔟 https://orcid.org/0000-0001-5289-2942

References

Andreassen, R. (2017). New kinships, new family formations and negotiations of intimacy via social media sites. *Journal of Gender Studies*, *26*, 361-371.

Barthes, R. (1980). *Camera Lucida: Reflections of photography*. London, England: Hill & Wang.

Berlant, L. (2008). *The female complaint*. London, England: Duke University Press.

boyd, d. (2008). Facebook's privacy trainwreck: Exposure, invasion and social convergence. *Convergence*, *14*, 13-20.

Caruso, A., & Roberts, S. (2017). Exploring constructions of masculinity on a men's body-positivity blog. *Journal of Sociology*, *54*, 627-646.

Dobson, A. S., Robards, R., & Carah, N. (2018). *Digital intimate publics and social media*. London, England: Palgrave Macmillan.

Friedman, M. (2018). Insta-judgement: Irony, authenticity and life writing in mothers' use of Instagram. *Interactions: Studies in Communication & Culture*, *9*, 169-181.

Gross, L. P., Katz, J. S., & Ruby, J. (2003). *Image ethics in the digital age*. Minneapolis, MN: University of Minnesota Press.

Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry*, *10*, 261-280.

Habermas, J. (1962). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge, UK: Polity.

Highfield, T., & Leaver, T. (2015). A methodology for mapping Instagram hashtags. *First Monday*, *20*, 1-11. Retrieved from https://firstmonday.org/article/view/5563/4195

Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*, 47-62.

Kanai, A. (2017). Girlfriendship and sameness: Affective belonging in a digital intimate public. *Journal of Gender Studies*, *26*, 293-306.

Locatelli, E. (2017). Images of breastfeeding on Instagram: Self-representation, publicness, and privacy management. *Social Media + Society*, *3*(2), 2056305117707190.

Markham, A., & Buchanan, E. (2012). Association for Internet Researchers' Ethics Working Committee—Ethical decision-making and Internet research: Version 2.0. *Association of Internet Researchers*. Retrieved from https://aoir.org/reports/ethics2.pdf

Markham, A. N., Tiidenberg, K., & Herman, A. (2018). Ethics as methods: Doing ethics in the era of big data research—introduction. *Social Media + Society*, *4*(3), 2056305118784502.

Marwick, A. E., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*, 114-133.

Morgan, D. (2011). *Rethinking family practices*. Basingstoke, UK: Palgrave Macmillan.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton, NJ: Princeton University Press.

Sennett, R. (1977). *The fall of public man*. New York, NY: Knopf.

Silverman, D. (2007). *A very short, fairly interesting and reasonably cheap book about qualitative research*. London, England: Sage.

Tiidenberg, K. (2018a). Visibly ageing femininities: Women's visual discourses of being over-40 and over-50 on Instagram. *Feminist Media Studies*, *18*, 61-76.

Tiidenberg, K. (2018b). Ethics in digital research. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 466-479). London, England: Sage.

Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, *12*, 313-325.

Author Biographies

Signe Ravn is senior lecturer in Sociology. Her research interests include research ethics in qualitative research with vulnerable youth. She is the lead investigator and contributed in the research design, data collection, data analysis, and the writing of this article.

Ashley Barnwell is lecturer in Sociology. Her research interests include social theories of ethics and representation, and life narrative. She contributed in the research design, data collection, data analysis, and the writing of this article.

Barbara Barbosa Neves is lecturer in Sociology. Her research focuses on the design and evaluation of digital technologies to enhance social connectedness among older adults, as well as on the ethics of participatory research. She contributed in the research design, data collection, data analysis, and the writing of this article.