Insights on Privacy and Ethics from the Web's Most Prolific Storytellers

Christopher Wienberg and Andrew S. Gordon
Institute for Creative Technologies
University of Southern California
12015 Waterfront Drive
Los Angeles, CA 90094
{cwienberg,gordon}@ict.usc.edu

ABSTRACT

An analysis of narratives in English-language weblogs reveals a unique population of individuals who post personal stories with extraordinarily high frequency over extremely long periods of time. This population includes people who have posted personal narratives everyday for more than eight years. In this paper we describe our investigation of this interesting subset of web users, where we conducted ethnographic, face-to-face interviews with a sample of these bloggers ($n = 11$). Our findings shed light on a culture of public documentation of private life, and provide insight into these bloggers' motivations, interactions with their readers, honesty, and thoughts on research that utilizes their data. We discuss the ethical implications for researchers working with web data, and speak to the relationship between large social media datasets and the real people behind them.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*ethics, privacy*; K.7.4 [**The Computing Profession**]: Professional Ethics—*codes of good practice, ethical dilemmas*

General Terms

Human Factors

Keywords

Privacy, Weblogs, Research Ethics, Human Subjects Research, Ethnography

1. INTRODUCTION

A very common paradigm for research on weblogs and social media involves algorithmic analysis of extremely large datasets. These analyses often mirror those used in other fields of science, where the qualities of and interactions between data points are the focus of investigation. However, social media datasets are distinct in that their data points describe the behavior and attributes of real people. Although the techniques of analysis are often shared, the human factor in these datasets raise unique concerns and challenges. In many ways, social media investigations can be viewed as scaled-up versions of the traditional human subjects research, where the aim is generalizable knowledge about people. Work on social influence [1], information diffusion and spread [7], and user behaviors [16] are fundamentally large-scale analyses of people cast instead as analyses of generic large datasets. Social media, precisely because it is mediated by the web, encodes the richness of human life in ways that lend human lives to algorithmic analysis. However, this mediation separates researchers from their true subjects, prohibiting investigators from understanding the larger picture of how this data is generated. What motivates these people to participate in social media, and to do so publicly, in the first place? How does a person's usage of social media relate to underlying characteristics of that person? How do users decide what to publicize and what to keep private? What are their expectations of their audience, and how should these expectations impact methodological considerations in social media research?

In our work we seek to provide insight into the people who are studied but rarely acknowledged in the course of large-scale social media research, with a focus on how these users perceive their audience and the repercussions for online research ethics. One approach to investigate this research subject is to conduct large-scale surveys, canvasing thousands of people for their opinions. This is the approach of the Pew Research Center, for instance, who regularly conducts mass surveys of web users in their Pew Internet & American Life Project [4]. A similar study correlated social media content to the personalities of its authors, as measured using various personality tests (e.g. the Big Five) [20]. This approach allows one to study a large breadth of opinions and behaviors. However, in our work, we adopt a more personal approach, undertaking the unusual step for social media researchers of meeting and interviewing several social media users.

In a survey of weblogs, Munson and Resnick [10] tell us that the typical weblog takes the form of a personal journal, read by a small group of friends and family. In these personal journals, weblog authors include personal narratives of their own life experiences among other content. Swanson [15] indicated that the fraction of these personal narratives is very small, estimating that only 5.4% of all non-spam English-language weblog posts are personal stories, defined

WebSci '15, June 28 - July 01, 2015, Oxford, United Kingdom

ACM 978-1-4503-3672-7/15/06…$15.00
DOI: http://dx.doi.org/10.1145/2786451.2786474

as non-fictional narrative discourse that describes a specific series of causally related events in the past, spanning a period of time of minutes, hours, or days, where the storyteller or a close associate is among the participants. Gordon and Swanson [5] have shown that these personal stories can be identified automatically. By employing supervised machine learning techniques, Gordon and Swanson identified one million of personal stories in weblogs in the ICWSM 2009 Spinn3r weblog dataset [3]. Wienberg et al. [19] noted that there was a high variance in the frequency of posts by personal story bloggers, with tens of thousands of people posting at a frequency between weekly and daily. The most prolific authors write almost every day, providing a glimpse into their day-to-day lives to any interested web reader.

In this paper, we focused on this particular subset of social media users, prolific weblog storytellers. We conduct unstructured, personal interviews with eleven of these prolific social media users, addressing three issues important to social media researchers: What motivates people to use social media, especially publicly? How honest are they in their usage, how well does their social media content reflect their lives, and how do they decide what to keep private? What do they expect of their audience, and what implications does this have for researchers?

2. RELATED WORK

The questions discussed this paper have have been discussed in previous papers before. Nardi et al. [11] provides a detailed ethnographic examination of the motivations of bloggers. They suggest that the reasons for blogging are as disparate as the content of the blogs. The most common motivations include: to document the author's life to readers; to use as a convenient platform to express opinions; as catharsis; as a muse to help the author form and sort through thoughts; or, as a tool to communicate with a community, citing the examples of a group of poets and an educational community. Nardi et al. also discussed blogging practices, including frequency (varied from once a month to multiple times per day) and their relationship with their readers. On this point, their description is sparse but poignant. It is clear that bloggers are careful about the amount of personal disclosure on their blogs: bloggers want to be open and honest, even when expressing controversial opinions, but not at the cost of strained relationships with readers—especially real-world personal relationships. As examples, the authors mention a blogger who tempers her political opinions to avoid irritating an uncle, another who was warned by his mother to temper his use of adult language because his grandmother reads his blog, and one who took down and replaced his blog with one he advertised less after accidentally hurting the feelings of a friend.

These issues are especially important in light of evidence that web users frequently post content they regret [20]. Most frequently these regrets relate to the user's audience (e.g. offending a reader or posting something that reflects poorly on the user). Stefanone and Jang [14]'s study on the use of weblogs to maintain relationships sheds some light on the relationship of bloggers and their audiences. They conclude bloggers sacrifice control of personal information to better communicate with friends and family, but caution that bloggers have questionable understanding of the full extent of their audiences and the true costs of disclosure. Similarly, in a paper on conducting adolescent medical research using social media (i.e. MySpace), Moreno et al. [9] simultaneously argue that web users who set profiles to public cannot reasonably complain when a researcher accesses their data without permission, but that it is at best unclear whether users understand this or intend to become research subjects.

Several studies have attempted to clarify users' understanding of their audience and what privacy means to them. One of the most well known studied the online behavior of teens, using ethnographic techniques to better understand how they navigate privacy in the online social web [2]. boyd observes that teens are developing new social norms and practices to establish privacy in a public space. boyd highlights techniques such as using privacy settings, disabling and re-enabling accounts, and using deliberately vague messages designed to obfuscate their meaning to outsiders. The most frequent outsider these techniques are intended to foil are parents and similar authority figures, and a major concern of nearly every person in boyd's study is parental intrusion. boyd's study participants indicate that their attempts to secure privacy frequently fall short, and show that unwanted intrusion is largely unappreciated by these web users.

Given the results of boyd's and other studies, caution seems prudent when conducting web research, even using public data. Unfortunately, research guidelines are evasive or silent—even deliberately so—on this matter. The US Department of Health and Human Services [13] and the Association of Internet Researchers [8] have both recently released reports on conducting internet research. Both take the stance that internet research is diverse and difficult to generalize over. Instead, these reports include open-ended questions and discussion about privacy and ethical issues, while steadfastly refusing to ground these discussions in concrete examples that may provide real guidance for researchers. These questions are certainly valuable in assessing research ethics and discuss crucial topics such as just when web data analysis becomes human subjects research, but without advice about how a researcher's answers relate to her ethical obligations these guidelines fall short of their goal of guiding researchers and Institutional Review Boards (IRBs) on how to conduct ethical internet research. We present this study as concrete example to discuss internet research ethics.

3. MEGABLOGGERS: PROLIFIC STORYTELLERS OF THE WEB

We began by identifying personal narratives on the web using the story classification technology developed by Gordon and Swanson [5], collecting a corpus of over 24 million personal stories posted to weblogs between 2010 and 2012. Using metadata associated with these stories, we identified weblogs that were active over the entire period of collection and where personal stories were posted at a rate of at least twice per week. From over 1.8 million distinct weblogs in our corpus, we identified 903 that fit these criteria.

We examined a sample ($n = 73$) of these personal weblogs with high activity. In keeping with early research on personal blogging, most (73%) of the bloggers were female [6, 12]. Interestingly, this selection of weblogs had a greater diversity of age compared to earlier findings indicating a strong skew toward younger bloggers [6, 12]. An age breakdown is presented in table 1.

Using this list of personal storytellers of the web, we sought to address three research questions. The first relates to the

Age Range	Percent
18–32	11.0%
33–43	39.7%
44–54	15.1%
55–64	16.4%
65+	17.8%

Table 1: Age breakdown of authors of sampled ($n = 73$) prolific personal bloggers.

motivations of these authors. Why do they blog in the first place? Why blog publicly about their personal lives? Why so frequently? These issues are important to research using personal social media data in general, and personal stories from prolific bloggers in particular, when using these data as a knowledge base for higher level reasoning tasks. These motivations will shape what these authors choose to write about and how representative it is of their daily lives, which is crucial when judging how personally invasive a research plan can be. The motivation to publicly share so much will help us gauge our ethical obligations.

The second research question we sought to answer relates to the honesty of these bloggers. Are the stories they tell actually drawn from their personal experiences or fabricated? What kind of framing or embellishment do they employ when writing? What kind of decisions do these authors make about what to publicize versus what aspects of their lives to keep private? These issues directly relate to the representativeness of the stories on the blog. If stories are greatly embellished, or outright fabricated, then their value as a knowledge source about real life is questionable. If some topics are systematically left out of the blog, then this should be considered when analyzing this data. If bloggers reliably self-filter their posts, then we can be more confident that analysis of their data is not unethically intrusive.

The third research question we had is with respect to these bloggers' understanding of their readers. Who do users expect are reading their blogs? Do they anticipate, despite posting publicly, they have some sense of anonymity and privacy because they have relative obscurity on the web? What do they think of researchers accessing and analyzing their content? If authors have no expectation that people are reading their blogs aside from a few close friends, then as researchers we should take pause when accessing, analyzing, and distributing their content.

To address these research questions, we conducted IRB-approved interviews with eleven prolific personal story bloggers[1]. These authors and descriptions of their weblogs are listed in table 2. Nine were conducted face-to-face in May 2013, while one was conducted by telephone in January 2014. Participants were recruited by direct email contact using contact information discovered on their weblogs, which we had very little trouble identifying—even for users who were careful to protect their identities. In recruiting participants, we were geographically limited by our decision to conduct face-to-face interviews, and therefore chose to contact bloggers in California, though the bloggers were spread across urban, suburban, and rural settings. The interviews were unstructured, allowing interesting related topics we did not anticipate to be discussed. The interviews were video recorded for later review. With advance permission from the participants, the video footage was edited into a short documentary presenting the work.

Due to decisions made in the process of contacting and interviewing participants, there are limitations to this work that are important to note. The small number of participants and the geographic homogeneity mean that the findings in this study are not generalizable to the entire population of prolific story bloggers. An immediately obvious repercussion of these limitations is the gender skew toward women (82% of participants versus 73% of prolific personal bloggers). Despite the inescapable limitations of the size and skew of our sample, these participants' views are valuable in that they provide a range of perspectives, and their remarkably similar practices and perspectives give us some confidence that our findings are informative, if not strongly generalizable. Additionally, the presence of video cameras during these interviews may have had a chilling effect on some of the participants responses. We are not very concerned about this effect for two reasons. First, the participants we spoke to were enthusiastic when we first contacted them, indicating that they were not concerned with the exposure. Second, we conducted a telephone interview with one participant who expressed interest in the study but did not want to be video recorded, to provide a perspective without the potential biases introduced by the camera. Her responses were largely similar to the other participants.

4. MOTIVATIONS OF PROLIFIC STORYTELLERS

The bloggers we spoke with have many reasons to start blogging. Many are motivated to tell others about a particular aspect of their lives. They develop themes around their blogs, where the stories they write are drawn from a particular facet of their lives. Three of the blogs we looked at had an explicit theme from conception: Meryl and Chris' blog about their home improvement projects, Tina's blog on finding loose change in the course of everyday life and drawing life-inspiration from these occurrences, and Daphne's blog on staying fashionable while being a mother.

Other bloggers we spoke to have a much looser theme to their blog. Frequently, these blogs started for a narrow purpose, and have expanded over time. Vicki started her blog as part of a research study for her daughter. Her daughter has Fragile X—a chromosomal condition—and her doctors asked her mother to maintain a log of her daughter's life. The blog started as this log, and just continued after the study was over. Similarly, Sally blogs about her daughter with disabilities. Sally started the blog for catharsis, after dealing with a particularly stressful challenge with an agency vital to her daughter's care. Now, she writes about her and her daughter's lives, focusing both on the travails of dealing with the various institutions and bureaucracies they come in contact with in dealing with her daughter's medical conditions and the stories of their everyday lives.

The remaining five bloggers are even more similar to a general diary or journal, though themes do emerge. Two bloggers—Lisa and Monica—fit the archetype of a *mommy blogger*, though at least Monica would take umbrage with that term. Monica's blog spun out of a scrapbooking project she and her husband undertook in the first year of her son's life. Interestingly, Lisa keeps a second blog. On it, she posts

[1] We sampled ten blogs, one of which is maintained jointly by a couple. They were interviewed together.

Blogger(s)	Description
Meryl and Chris	Jointly maintain a blog about their home improvement projects. Both participate in content creation, but Meryl primary writes and posts.
Tina	Writes about discovering loose change in daily life.
Daphne	Blogs about fashion and raising her children.
Vicki	Mostly posts photographs of her daughter, who has a chromosomal disorder.
Sally	Writes about her life and her daughter, who is disabled.
Lisa	Maintains two blogs: one about raising her young children, the other with more adult personal content (e.g. going to concerts or getting tattoos)
Monica	Blogs about family life with her husband and two sons.
Bo	Blogs about her personal life, which is spent running a farm.
David	Blogs about memories he recalls over the course of the day. David is a school teacher.
Sara	Writes about her personal life, frequently about work in the tech industry or starting a second career as a romance novelist.

Table 2: Descriptions of study participants and their weblogs.

more risqué content, such as going to concerts or getting tattoos. She decided that she wanted to share these experiences as well, but they did not fit her original blog due to both its typical content and its audience. Bo's posts, because she lives on a farm, naturally center on aspects of farm life like caring for livestock. Bo had been journaling for many years before she started blogging. David and Sara's blogs are the last blogs without an easily articulated theme, and were started as an opportunity to journal.

4.1 Capturing life, everyday

While Bo started her blog as an extension of her offline personal journaling—after being inspired by the film *Julie and Julia*, about a woman who blogs her attempt to cook all the recipes in Julia Child's *Mastering the Art of French Cooking*—she never expected it would grow into what it is today. At first she intended it for "when something really interesting happened." Instead, it is an almost daily enterprise, in part because "having to focus on something that you think is interesting makes you look at your world entirely differently." If she skips a day or two, she starts getting phone calls from concerned friends and family asking after her well-being.

Sara gives a similar reason for blogging so frequently:

> my parents can read the blog the next morning and they can know that I was alive at midnight last night and they don't have to call me to verify that I was alive.

Additionally, since so much of her extended family lives far from her, blogging often is "a good way for them to feel connected; then I only have to talk to them once a week." Not that she's trying to avoid her family; rather, the blog is a different, more efficient channel of communication that saves her family from worrying. Sara also cites that posting so frequently becomes habitual, and that when she has a streak of many days in a row with posts she "can't break this streak, and I need to just hammer something out."

David is also compelled by the streak. He has posted every day since May 2005. It is so important to him that, if he knows he will be without internet access, he will write and schedule content to be posted in his absence. He does this because it gives him a sense of fulfillment and he finds it important "to do things regularly [. . .] to keep a pattern."

Vicki and Lisa both blog frequently to keep friends and extended family current with the lives of their children. From the beginning, Lisa blogged almost every day. She was driven to keep distant family—especially her husband, who is often away from home for work—informed about her children's lives. She initially started this process by sending daily emails with photographs to close friends and family. Eventually—on her mother-in-law's suggestion—she thought, "maybe everyone's getting tired of getting flooded with pictures, so I decided to start a blog so that they could access it on their own."

Vicki, similarly to Bo and Sara, gets phone calls from friends and family if she misses a day. Because of the commitments of her daily life, Vicki cannot blog every day. However, since her "family wants to see something every single day," like David she will "sit down 2 or 3 times a week and I'll publish ahead 2 or 3 posts so that they'll get a new post every single day." This works well because she's a self-described "avid photographer," and every post has upwards of dozens of photographs of her daughter. Vicki takes so many pictures that her posts are actually several months behind reality, e.g. pictures of Christmas are posted in March, summer vacations are posted in October.

Daphne, Chris and Meryl blog so frequently out of a sense of obligation to their readers. Daphne says, in the past, she's received a message from a reader "that mentioned when they open up my blog and it's the same post, they are a bit disappointed." Meryl reports that when she and Chris started their blog they "didn't post as much." Now, she says:

> I want to give people something to read, so I feel a sense of wanting us to complete projects so I have something to write about, so that we can share, because people go to your blog to read it. [. . .] I feel like I'm doing them a disservice if I'm not posting.

Like Chris and Meryl, Sally never anticipated she would post as frequently as she does. Despite this, she says, "once I got this rolling there were so many stories of [her daughter's] life, they were just spilling out of me." She notes that she's slowing down, in part because she's exhausted a large part of the backlog of stories.

Tina blogs so frequently because of the the theme of it—discovering loose change and framing it in an inspirational, heartwarming way—necessitates this type of writing. She had a streak of over 1000 days of finding discarded coins and writing about the story and her take-away lesson from the day. While Tina was content to share these stories via email to friends and family, she was repeatedly encouraged "to put it on the blog so it could reach a greater audience."

4.2 A public forum

Many of the participants indicated that they started their blog as a communication channel to friends and family. Despite targeting a limited audience, they decided to make their blogs publicly accessible. Lisa indicated that she does not use a platform that is more private, such as Facebook, because of readers she wants but would be unable to reach that way. For instance, her grandparents and father "don't really use a computer, don't really have a Facebook, so in order for me to feel like I can keep them updated" she needs to use something easier for them.

Vicki speaks about having a few audience segments: her family, local moms she has connected with, and parents of children who share her daughter's medical condition. In order to reach audiences beyond her family blogging publicly becomes a necessity. She consciously decided that she would make it public, as a resource for other parents in similar circumstances. "It's free and available for people," she says. "It's tagged, they can research it. [. . .] I'm constantly referring people back to it."

Daphne likes the accountability of blogging publicly. By having an audience to appease, she has to keep up with her own goals. She says she keeps a private journal "on my feelings and emotions," but the blog lets her put her fashion opinions and work out there "like art" and get feedback from readers. Sara also keeps a private journal, but likes blogging because she can "get some recognition for what I write." She likes that it sets her apart from her friends, none of whom blog. "They're more like workers or consumers of content and not producers," she states about them.

Ultimately, the bloggers indicated that the point of the medium is to let others read it, so blogging in a non-public way would not make sense.

4.3 A long-term hobby

These authors not only stood out for the frequency of their posts, but also for how long they have been writing. When we asked these authors what might cause them to stop, we received many different answers. David has been asked by several people when he would stop. He said, "it occurred to me, 'I could just stop now." Nobody's paying me, I'm not getting any significant feedback from anyone saying, 'no don't stop.' " He keeps going because it has become a hobby that is important to him.

Sara and Sally each plan to blog as long as the content will carry them. Sara says:

> if I go through a week where I look back at my posts and read that I did nothing but get up, go to work, and come home and watch TV with my roommate, and go to bed, I'm like, why do I even bother and why does anyone read it?

Sally says "it is slowing down a little, but I don't see it stopping either." She's already worked through a large backlog of memories about her and her daughter, but still has current tales to keep writing, and occasionally remembers something from the past worth telling. The only definite reason Sally might stop is if something happened to her daughter, presumably related to one of her medical conditions.

Lisa expressed that she would stop if her children and family were in danger as a result of her blog. Presumably, the threat of danger would come as a result of content posted to the blog—Lisa expressed that the crime drama television program *CSI* makes her nervous about such a possibility. Vicki said she would stop "if my daughter asked me to." She said she would even take the blog down if asked by her daughter, given how much of her daughter's life is recorded on it. Even so, Vicki says she would probably carve out "a private area when I'm talking about her" and keep a public blog of things unrelated to her daughter.

One of the bloggers we spoke to, Tina, actually stopped for a little while. When interviewed, Tina expressed that she really disliked writing. The culmination of her blog was the building of a home she and her husband designed. The blog, and its associated hobby of finding loose change, helped her work through years of setbacks on building her dream home. Once her new home was built, Tina took a long hiatus from blogging, as a chance to recover, and because part of its initial drive was complete. She has since returned to it, adopting additional projects to write about.

Daphne, Meryl and Chris express a sense of obligation to their readers as a reason for continuing to blog. Additionally, Meryl says she and Chris find blogging really fun. On the horizon is a possible end date: when they are finally done renovating their home, though she added maybe they "would have another house that we could do so I could blog about it." Daphne is compelled by reader feedback. "I love gaining comments," she said. "It's great because it shows that someone is reading you."

As such a long running record, several of the participants think of their blogs as a nice resource documenting their lives. Chris and Meryl say that blogging is "a nice way to remember your own life." Vicki thinks it could be a great resource for her daughter, when she is older, to look back nostalgically. Sara actually finds the blog to be such a helpful resource for looking back on her life that it's one of the reasons she keeps blogging:

> If anyone in my friend group has a dispute of when something happened, we just search my blog for, like, words that may pop up on the day it happened.

Also, friends who have moved away can keep up on her life and when they see each other again "they're like, oh we know what you did." Monica likes the ability to look back at her kids' lives and thinks of her blog as a public scrapbook.

In order to keep such a consistent throughput over an extended period of time, many of the bloggers indicated they have a routine to writing. David and Bo write every morning before going to work or tending to the farm, respectively. Sara writes at the end of the day, just before going to bed. Lisa blogs at night after her children go to bed, so she can give it her concentration. Other bloggers had other routines that fit their schedules which helps them maintain such consistently frequent blogging.

Interestingly, each of these bloggers have taken this obligation without hope or expectation of financial gain. Sara even expressed incredulity at the popularity of the genre of blogging she is a part of, saying "it seems like a lot of people aren't blogging in this way anymore [. . .] they're all trying to build a platform." Only one blogger reported ever having made any money from the blog, Lisa:

> I did the advertisement one time, [. . .] they paid

> me like twenty dollars for it, but I didn't really care about the money, it's always been for the benefit of our family.

Monica indicated she received press tickets to live musical theater and child-care products on a few occasions, so that she would write about them in her blog. While she accepted a few things, ultimately she decided it was neither enjoyable nor worth her time. Bo, Chris and Meryl all said it would be nice to profit from their blog, but have never tried. Meryl indicated she and Chris would not compromise their integrity, saying, "I wouldn't want to ever advertise a product that I wouldn't use myself." Vicki said she would never take the path of monetization that seems easiest to her: turning her Fragile X documentation into a book. Her mother has been encouraging her to do just that, but Vicki says she is not interested because a book "wouldn't be so interactive, and it wouldn't be searchable, and it wouldn't be as dynamic."

5. HONESTY AND FRAMING IN PUBLIC STORYTELLING

An important question for these bloggers relates to how well their blogs represent their offline lives. Currently, when collecting stories, the only account of an event we have is the story a blogger has chosen to write. If a blogger is being dishonest—omitting important details, spinning the event in a more favorable light, or outright fabricating the story—then this could have repercussions depending on how that information is analyzed and used.

Bloggers tended to frame their stories in a more positive light. Vicki says:

> I sort of have this lens about what I write. I do sometimes write about deeper things, I write about her development, I write about struggles, but I write about them in a vanilla sort of way, so I'm not seeming urgent or I'm not worrying anybody in my family.

Similarly, Sally states, "one thing I do on the blog, I give things a little more of a positive spin." She continues, talking about the challenges and fears resulting from her daughter's medical conditions:

> I've put some pretty deep, pretty scary stuff on there about her medical stuff and where we've been. You know being on the edge of a cliff because that's the image I have when she's come close, we've come close to losing her many times and to me it's a cliff. And people who read my blog a lot will know that, that we live sometimes two inches from the cliff, and that's the safe zone. We are never very far from the cliff. But I have been very honest about that stuff, but I still frame it in a more positive way than, sometimes than I feel. It's not a lie, but it's purposeful. To put a happier face on this, not unrealistic. [. . .] Sometimes it's automatic. And sometimes it's, 'OK, I need to make this a little happier, more positive,' you know? But I'm not lying.

Trying to keep the blog toward the positive was common among the bloggers. Monica likes to think of her blog as a personal resource to "capture these little moments I don't want to forget [. . .] tiny things I would never remember again," rather than a log of her entire life. She sees no point in recording the bad days unless there's a lesson to be learned, because when she returns to memories of past days, it is much nicer for her to see the little, happy moments.

Meryl indicated that she and Chris actively post stories that cast them in a negative light. She says:

> I would say I avoid painting things as much more rosy than they are. Like a lot of people who are couples won't blog about things they get in a fight about, but it's totally natural [. . .] it's important to me that you admit when things don't go according to plan or fail, and that's prevalent I think for us.

A few bloggers we spoke to cited narrative improvement as a reason they might alter the stories the post. David confesses to a little lying for "dramatic effect." Sally admits she edits for narrative improvement. "Different details that change the feeling of it," she says, and so she will switch around narrative order in order to compose a good narrative.

In keeping with findings in other forms of social media [17], a few of the bloggers reported having posted content they later regretted. Lisa indicated she once regretted a post about a fight she had with her husband. She says:

> he was really upset that I had, like, publicly written about it [. . .] I was just mad, and felt like I needed to write about it, and wasn't really thinking about how other people would perceive him whenever I did that.

As a result of that experience, Lisa states, "since then I've made sure not to say anything that I think would upset him or anyone else really."

Sara also described an occasion when she regretted something she posted. Before her current blog, she maintained a previous one, which "sort of died this fiery death." There were a group of people she knew who were:

> passive aggressively blogging about, sort of, issues that were happening with our friends or within our friend circle, which, I don't know, I think I learned a lot early about what I wanted to post and what I didn't. You know, early on, I was treating it a lot more like a journal, and I was blogging pretty honestly about my feelings and what was happening with other people

Also, Sara indicated that she has "a period of six months that's now missing" from her blog. She says that there was "this guy that I was infatuated with, and I was posting very sort of depressed and sad things about, because he didn't like me back." After a while of this, she realized:

> this is not how I want to portray myself, and I don't think it's healthy to sort of keep dwelling on it, so it was kind of like a hard reset.

Now she says she has learned from these experiences:

> I tend to filter before I ever type. You know, so if, for example, if I had told a friend that I was, I couldn't go out with her because I was busy that night, I was going to stay in and work, and then I ended up going out with somebody else, I will sometimes just omit that I went out with the other person on my blog, so that the other person doesn't know that I basically preferred going out with that other person.

David avoids posting about "war stories" from his past, for the sake of his family:

> There was a period back where I was doing a confessionary about my drinking habits and so on. It was always something that my family really enjoyed not talking about, especially when it had gone and things were happy again. It's definitely been put out there, it's the reason I moved [...], got married, and had a son. If I was still drinking and taking drugs, I probably wouldn't be here and so for me it's a happy story. There are a few stories over here that some of my family members would say, "let's not mention that."

On this, and other subjects, David says:

> I've learned over time that there are certain subjects to kind of avoid [...] If I notice a bare wire there or there I'm like , 'you know we can sorta drive around that, we don't need to jump up and down.'

Other bloggers indicated that they have been highly conscious of what they are posting from the beginning, and take particular care to avoid posting things they would regret. As Sara indicated, she has learned over time to choose wisely what she posts. Monica says she has not posted anything she has regretted yet, in part because she very strictly avoids topics that are likely to cause controversy. She says, because posting about these topics require a lot of time and effort to choose one's words carefully enough to avoid upsetting people, she just avoids them. She also avoids posting things that are "too personal," and specifically mentioned a hard rule against posting photographs of other parents' children without their permission. Meryl and Chris indicated that they avoid writing about emotionally charged content, and Chris says about his personal blog, "once my grandmother started reading it, and nitpicking my grammar and that kind of stuff, I'm like, maybe I'll talk a little less about girls, or whatever." Bo, Sally and Tina indicate that they avoid using real names. Sally takes particular care when posting about others in a negative context:

> I will criticize the various agencies, but I won't criticize a person at an agency [...] because I feel like they're doing their job, it might be a crappy job. They might be in a crappy job and doing it the best that they can and doing it with the best intentions, but that [calling them out by name] is crossing the line.

Every blogger indicated that they believe they are being honest on the blog. The bloggers admit to using embellishments, elisions, and other techniques to modify their stories, but in their view, these techniques do not affect the overall honesty of those stories.

6. READERS AND INFORMATION CONTROL

Many of the bloggers we spoke to indicated that they started their blog for a few friends and family members to read. Lisa says she did not plan on having strangers reading her blog. "It just sort of happened," she says. She continues:

> I put up like a counter that told me where people were looking in on it from, and I thought it was pretty cool to see people from other countries looking at it and stuff like that.

Lisa suspects that people find her blog looking for vacation ideas, since her hometown is a popular vacation destination. Sara said her blog was always "targeted at people I actually know." That said, she continues, "I really just don't care if strangers read my blog, it's just not my target audience."

Every single blogger indicated that they knew they had strangers reading their blog, though often they had no idea how those readers found it. A common refrain when we contacted bloggers was, 'how did you find my blog?' Especially for those bloggers who have not been actively recruiting readers, these bloggers frequently wondered how perfect strangers found their blogs. David speculates:

> I don't know if you ever play Blogger Roulette, it says "Next Blog." And sometimes when I finish I want to know what the rest of the world is doing today and I'll click "Next Blog." [...] There are a lot of things, and again the Internet is out there just waiting to be looked at.

Chris and Meryl think that people are finding them from search engines. They say that they do not actively recruit strangers to read their blog:

> My dad tells me all the time [...] about things I should do to get more visitors, or what stuff I should look into, and I don't understand it, and I don't really care that much. [...] one day I just started [our blog], and somehow someone found it to start, like, no one was sharing it, now some blogs share it

Monica started her blog as part of a larger community. When she started blogging, the mommy blogger phenomenon was just getting started. "I didn't think anyone would really read at first," she says. Through her blog she made a lot of friends online. When she started, there was also an online community that she plugged into, including some now-famous bloggers. She says she corresponded with them, even though she had never met them.

Tina actively recruited readers in face-to-face communication. She would print up business cards with the web address of her blog, and use finding loose change as a conversation starter to get people interested:

> Perhaps I'd been in a store [...] where I have found a coin and I have picked it up in front of them, and I generally give them one of my penny cards, and say, thank you for sharing a smile with me, or whatever the comment is appropriate to them, and from there they will log on, tell other people about it.

Many of these bloggers indicated that they have connected in-person with strangers who read their blogs. Monica met many people she had been corresponding with online when she attended a conference for female bloggers. Meryl and Chris have met a few readers. One was someone local to them who sought them out on a walking tour of their neighborhood. Others they met on a renovation road trip they went on and blogged about. The son of the owner of a long-closed restaurant that David frequented as a child reached out to David after a post about the restaurant. Sally says she frequently meets new people reading her blog, and is always surprised by how many new readers she meets. The blog is required reading at a medical school in the city she lives so there are always strangers reading it. Daphne reported she has received a package from a reader in the past.

None of the bloggers reported they received contact from strangers that made them uncomfortable. However, because they are posting publicly, most of these bloggers have utilized some techniques to obscure their identities or activities. "I don't usually blog in advance what I'm doing [...] and that's partially a safety thing," says Sara. "I think every writer I know has had, maybe not a stalker, [...] you know one fan who is a little too *Misery*-ish for comfort." Despite this fear, Sara says, "anyone who read for more than a couple of weeks would start to pick, like, these are my favorite cafes, this is like what I tend to do." Lisa indicated she also does not post about what she is doing, but what she has already done, "so no one can be, like, 'oh, they're going to be at this location on this day.' "

Bo and Monica both indicated they don't use real names of people, in part for this reason. Monica says:

> If anybody actually tried, they could figure out who I was online, it's not very hard. But what I didn't want was for someone to be able to Google my kids' names and have all the stuff pop up

This affords her family some anonymity, even in public. People who are familiar with Monica's blog know or could fairly easily determine who is being described, where Monica's stories are set, and other details about her family's life. She also says blogging is getting "more complicated with my kids' privacy," and is posting less often about them as they age.

A notable technique a few bloggers used was some segregation of information. Sara and Lisa both reported having a second blog. Each had a particular audience in mind for the second blog. Lisa started hers for certain content:

> all the things I thought my dad wouldn't necessarily be interested in, and he's a pastor, so, anything he wouldn't want me to talk about in church, then that goes on that blog. [...] mostly there's the concerts, and then, like whenever I got my last tattoo, I put that on there, because I never really talked to my dad about tattoos, so I wasn't sure how he'd feel about it; now he knows that I have it. Just that, and I go with my mother-in-law to a fairy festival every year, which is far from being within his religious views, so in the beginning I put pictures of that up, but since then we've talked about it and he's ok with it, so I'll still put up pictures on my other blog, it's not really a big deal.

The second blog provides a place for Lisa to post more details about her life, while not directly exposing all readers to that content.

Sara used her second blog in the converse manner, giving a wider audience access to some content, but not providing them access to her full personal blog. She says:

> I went to India for six months with one of my old jobs, and that one I made a separate India blog, that was just that 6-month period. So I stopped posting on my personal blog, I blogged on the India blog for 6 months, and the idea was at the time I didn't want my coworkers to know about my personal blog, but if I set up this India blog, I could give them a link and they could see that without having access to the rest of it.

Given all these techniques that these bloggers use to obscure their identities, facts about their lives, and to compartmentalizing information to different audiences, we wondered about these bloggers' expectations about researchers accessing their content. Nearly every blogger was surprised by academic interest in their social media use. Monica, who works in the tech industry, said the possibility of academics analyzing her blog "never occurred to me." While all of them knew that strangers were accessing their blogs, there was no direct recognition of the implication that researchers may be analyzing their behaviors and content.

When presented with this possibility the bloggers had few concerns. They had already made a conscious decision to post publicly and accepted that their content could be accessed and used in ways they never anticipated. As Sally puts it, "I don't think that anything you put on the internet is private at all. Ever. I don't care what anybody says. And anybody that thinks opposite I think is crazy." While these users seem to have accepted that they have surrendered a lot of control of their content by posting publicly, they still had some wishes about how it might be used. Daphne expressed a limit on how she might want others—not researchers specifically—to utilize her content, saying:

> I don't write anything I wouldn't want XYZ of the population to read. The photos, I definitely would be more concerned about and wouldn't want people using them without my consent.

Chris and Meryl both indicated that they would want to know if their content was used in academic research. However, Chris posits, "ethically do they [researchers] have an obligation to [inform users]? No, not really." While Bo mostly treated researchers accessing and using her content as a given, she did say, "I'd like to know how my information is being used." She is of the view that:

> anybody that goes on any computer hopefully realizes that privacy in that venue is an illusion. So when you choose, or when I chose to 'go public,' I recognized that that was a danger.

Vicki indicated that she is very open to researchers accessing her blog. She says she is a "very open person" and is very comfortable with research, given her experiences with her daughter's medical condition. She recognizes, however, that she—by virtue of her openness—may not be representative of the population as a whole:

> even though I'm OK with it, I don't think everyone is since I think there should be some level of protection I guess. But I don't know how we would do that.

In light of having met researchers accessing their content, and having been informed of an entire field of research interested in social media, none of these bloggers indicated that they would change their behavior. Monica—who was hesitant to participate initially, and declined to meet face-to-face, citing her privacy—states she will not change her use of social media "because I really do think of writing as like a newspaper column."

More generally, these users acknowledged that analyses of user data are occurring all the time, and that ultimately it is unavoidable in the digital age. Meryl acknowledged that product targeting is happening all the time. "Amazon knew we were pregnant before our parents did," she says. Living in this world, she takes the respite she has, saying, "I always check the box that like, don't sell my information." Vicki says she sees great benefit from the analysis of her data:

> So I actually think its pretty cool that people are out there collecting data on us to be able to market directly to your needs and wants. And I think some people think that's an invasion of privacy and don't like that, but I would much rather have access to my media and have where I'm being advertised in the future and the advertising that I'm getting be something that I care about and am interested in.

These bloggers are operating on the assumption that, by posting on the web, information about them could be accessed and used in many ways they never imagined. They treat this as the trade-off of being able to reach a wide audience, and accept researchers and marketers accessing their information—even if, for some, they would prefer otherwise.

7. DISCUSSION

In this work, we've examined a subculture of people who decided to share widely and often about their personal lives. They have many motivations for doing so. Some are trying to be helpful, or to educate a larger audience about an issue about which they are knowledgable. Others wanted to communicate with particular people, and settled on a public weblog as the best means to do that. Still more saw public blogging as an opportunity to be creative and to reach a potentially wide audience. For the most part, they never expected they would have more than a handful of readers.

Now that they have been blogging like this, many of the people we spoke to feel a connection to their audience. Some never planned to write as often as they do, but now they feel as if they will be letting their readers down if they slow or stop. To others, having an audience gives their blog purpose: it inspires them to go on, it makes them feel as if they are having a positive contribution on the world.

We discovered that most personal storytellers of the web believe they are honest, and almost all dishonesties are by omission. Whether it is to omit negative experiences, or to remove parts of their day to obscure their activity, or just because its not that interesting, these authors do not share everything about themselves. That said they do share a lot, and having both read their blogs and met these authors, it is easy to construct a fairly accurate mental picture of these people from their writing. Interestingly, these bloggers have no incentives to post other than to share their life experiences with others. None of these bloggers reported making more than a nominal amount of money or receiving gifts on more than one or two occasions. For these authors, the goal is to supply readers with interesting content, with no profit incentives. This leads us to conclude that the personal stories of prolific personal bloggers are a good data source for research into their personal lives, with a few caveats. First, while finding stories of people describing their own socially questionable behavior is possible [18], these topics will be poorly represented by users who aim to keep their content positive. Additionally, analyses that depend on personal details at least as intimate as event participants' names are likely to fail, due to users' propensity to omit or obfuscate these details. Despite these caveats, when analyzing personal events or details outside these domains it is our estimation that the personal stories of the most prolific bloggers are a rich source of information about everyday life.

While these authors could be characterized as very public people, most of them do their best to exercise some control of how information about themselves are spread across the web. Whether by avoiding using real names of people and places, posting about past events rather than future plans, or even compartmentalize their posts across different weblogs, these authors are savvy and make conscious choices about how they share information about their lives. These authors freely admit that their techniques are mostly shallow barriers. Any determined adversary could uncover a significant amount of personally identifying information for each of these bloggers. These bloggers are making trade-offs, balancing a public platform with the preservation of some of their own privacy and control of their information.

Interestingly, while there was diversity in the kinds and quantities of information that the people we spoke to are willing to share, this diversity is very narrowly bounded. There were many more commonalities of practice between these authors than not. We found that the savviest users are careful in what they share, naturally filter details they feel are negative, and have developed a clear personal policy about how they interact with social media. Barring a fundamental shift in social media use, we expect this practitioner's perspective is what all social media users will converge to as they become more comfortable in the new digital landscape.

The balancing act these authors perform, between sharing widely and protecting their privacy, is a concession that web users ultimately have little control over their privacy. It was surprising that nearly every blogger believed that everything they post is likely to be known by someone they do not know. They operated on the assumption that anything they do on the web, including ostensibly more private activities than blogging like writing email, can be monitored by any number of people and organizations. For context, we conducted these studies before the recent revelations about the breadth and depth of US domestic surveillance programs, meaning distrust of the web with respect to privacy goes beyond worries about espionage and signals intelligence.

Our findings compel us to make a few recommendations with respect to conducting ethical research using social media data. It needs to be highlighted that social media users are not aware of internet and social media research. A strong case can be made that these social media users are among the savviest of the web: they have a strong understanding that their data can have a broad, inestimable reach, and have tailored their practices to reflect this. Given that even these users were very surprised to learn that academic researchers were interested in their blogs, the common claim that using a public profile serves as tacit consent to researchers should be dismissed on face. This is not to say that public social media profiles should be off limits to researchers, even without informed consent. While the ethical guidelines mentioned earlier in this paper are lacking in concrete recommendations, they are reasonable in their arguments that new paradigms in research ethics are necessary for the digital age, including a divergence from old principles regarding informed consent. That these participants were enthusiastic about the prospect of researchers studying their profiles is indicative that web users can be comfortable with being unwitting research participants, even if they are unaware of the prospect. We temper this claim by again noting the high degree of experience these users have in dealing with issues of disclosure and public sharing. These users have developed practices of sharing over the course of years of posting occasional mistakes. They are very careful now, and know their limits with respect to the sort of content they post. When dealing with less experienced users or with more sensitive content, researchers should be particularly concerned with the people behind these social media profiles. This can include taking great care to ensure data is anonymized before being shared (or not shared at all) to minimize potential harm to participants, or going as far as to contact users to gain informed consent. As data becomes more personal and intimate, large scale web researchers become less like data scientists analyzing a large dataset and more like social scientists or psychologists conducting a human study.

As our final point, we would like to urge the importance of considering the people generating social media. As datasets get larger it becomes easier to lose sight that, ultimately, these datasets contain a glimpse into the personal lives of thousands or even millions of people. Researchers should take care to think about the people who they are working with, indirectly, when they analyze social media content and behavior. They should keep in mind the motivations of these users and how well they represent themselves with their social media content and behaviors. Most importantly, researchers should always protect research participants, whether the research is done face-to-face or by analyzing the web.

8. ACKNOWLEDGMENTS

The projects or efforts depicted were or are sponsored by the U. S. Army. The content or information presented does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

We would like to thank Bo, Chris, Daphne, David, Lisa, Meryl, Monica, Sally, Sara, Tina, and Vicki for opening their homes and giving their time to participate in this study.

9. REFERENCES

[1] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, 2011.

[2] d. boyd and A. Marwick. Social privacy in networked publics: Teens attitudes, practices, and strategies. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, Oxford, United Kingdom, September 2011.

[3] K. Burton, A. Java, and I. Soboroff. The icwsm 2009 spinn3r dataset. In *Proceedings of the Third Annual Conference on Weblogs and Social Media*, ICWSM 2009, San Jose, CA, 2009.

[4] M. Duggan and A. Smith. Social media update 2013. Technical report, Pew Research Center, 2013.

[5] A. Gordon and R. Swanson. Identifying personal stories in millions of weblog entries. In *Proceedings of the Third International Conference on Weblogs and Social Media, Data Challenge Workshop*, ICWSM 2009, San Jose, CA, 2009.

[6] S. C. Herring, I. Kouper, L. A. Scheidt, and E. L. Wright. Women and children last: The discursive construction of weblogs. In *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*, 2004.

[7] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on digg and twitter social networks. In *Proceedings of the Fourth International Conference on Weblogs and Social Media*, ICWSM 2010, Washington, DC, 2010.

[8] A. Markham and E. Buchannan. Ethical decision-making and internet research: Recommendations from the aoir ethics working committee (version 2.0). `http://aoir.org/reports/ethics2.pdf`, 2012. Accessed: 2014-11-10.

[9] M. A. Moreno, N. C. Fost, and D. A. Christakis. Research ethics in the myspace era. *Pediatrics*, 121(1):157–161, 2008.

[10] S. A. Munson and P. Resnick. The prevalence of political discourse in non-political blogs. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, ICWSM 2011, 2011.

[11] B. A. Nardi, D. J. Schiano, M. Gumbrecht, and L. Swartz. Why we blog. *Communications of the ACM*, 47(12):41–46, 2004.

[12] S. Nowson and J. Oberlander. The identity of bloggers: Openness and gender in personal weblogs. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pages 163–167, 2006.

[13] Secretary's Advisory Committee on Human Research Protections (SACHRP). Considerations and recommendations concerning internet research and human subjects research regulations. US Department of Health & Human Services, March 2013.

[14] M. A. Stefanone and C.-Y. Jang. Writing for friends and family: The interpersonal nature of blogs. *Journal of Computer-Mediated Communication*, 13(1):123–140, 2007.

[15] R. Swanson. *Enabling Open Domain Interactive Storytelling Using a Data-Driven Case-Based Approach.* Ph.D. dissertation, University of Southern California, 2010.

[16] C. Wagner, M. Rowe, M. Strohmaier, and H. Alani. What catches your attention? an empirical study of attention patterns in community forums. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*, ICWSM 2012, Dublin, Ireland, 2012.

[17] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. "i regretted the minute i pressed share": a qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, Pittsburgh, PA, 2011.

[18] C. Wienberg and A. Gordon. Privacy considerations for public storytelling. In *The Eighth International AAAI Conference on Weblogs and Social Media*, ICWSM 2014, Ann Arbor, Michigan, USA, June 2014.

[19] C. Wienberg, M. Roemmele, and A. Gordon. Content-based similarity measures of weblog authors. In *The 5th Annual ACM Web Science Conference*, WebSci'13, Paris, France, May 2013.

[20] T. Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373, 2010.