Social Media Ownership: Using Twitter as a Window onto Current Attitudes and Beliefs

Catherine C. Marshall
Microsoft Research, Silicon Valley
1065 La Avenida
Mountain View, CA 94043
cathymar@microsoft.com

Frank M. Shipman
Department of Computer Science
Texas A&M University
College Station, TX 77843-3112
shipman@cs.tamu.edu

ABSTRACT
Social media, by its very nature, introduces questions about ownership. Ownership comes into play most crucially when we investigate how social media is saved or archived; how it is reused; and whether it can be removed or deleted. We investigate these social media ownership issues using a Mechanical Turk survey of Twitter users; the survey uses open-ended questions and statements of belief about realistic Twitter-based scenarios to give us a window onto current attitudes and beliefs. Our findings reveal that respondents take a liberal attitude toward saving and storing the tweets that they encounter. More caution is exercised with republishing the material, and still more with sharing the material among friends and associates. Respondents approach removal of this type of lightweight social media most cautiously. The material's provenance and the respondents' relationship to the material (whether they are the author or subject) has considerable bearing on what they feel they can do with it.

Author Keywords
Twitter, social media, information rights, survey, reuse.

ACM Classification Keywords
H4.3. Information Systems: Communications Applications.

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION
Services such as Facebook, Twitter, and Flickr introduce new questions about the ownership and control of social media [16]. Although legal decisions, license agreements, and explicit policies address these issues, the people who use social media services may not be aware of them, nor adhere to them. Because social media may be created jointly and may document or record events involving many people [20,22], it presents design challenges for new digital services such as archiving or publishing.

What do people feel they can (and cannot) do with information created by others? We divide potential activities into four basic categories: (1) *saving* digital content; (2) *sharing* digital content with specified (and limited) groups of people; (3) *publishing* digital content so it is broadly accessible; and (4) *removing* digital content from its social media venue (and not necessarily deleting it from local storage, in line with reported behavior [21]).

CSCW applications have long demonstrated that users may not apply reciprocal standards to actions they feel they can take [5]. For example, they may want to see others without being seen, or they may want to have capabilities that are not offered to their peers. In this case, we investigate what respondents feel they should be able to do with microblog content they encounter, as well as probe what respondents feel that others can do with the respondent's content.

Finally, we are interested in scaling up our findings from individuals to public institutions: what do people feel that public institutions should be able to do with today's ephemera and everyday digital belongings? For example, some Twitter users were unhappy when the Library of Congress announced plans to archive public tweets; they felt strongly that the service offered them privacy through obscurity. How common is this sentiment?

Social media encompasses a number of different forms and genres. This study is the first in a series of studies we have designed to investigate peoples' attitudes (and in some cases, their reported practices) about social media ownership, archiving, and reuse. Because different types of social media have different properties—for example, emotional heft, creation ease, file size, and representational fidelity—we are limiting each study in the series to involve a single social media type.

This study is based on content from Twitter, a microblogging service in which users share short messages (140 characters or fewer) with a set of subscribers (although anyone may read the messages if an account's privacy settings allow them to). The messages, or tweets, may stand alone or they may be links to published material on the

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Web (photos, blog posts, articles, etc.), or to content stored in sister services such as Twitpic.

If we are ultimately interested in persistent media, why look at Twitter first? Twitter tweets are largely categorized as lightweight bits of communication, and as such, are sometimes considered ephemeral. However, if we look more closely, tweets walk an interesting line between epigrams and microblog posts to news flashes and ambient pop culture material that is just 'in the air' at a particular time or place [3,8]. Yet people do save and reuse Twitter tweets (both their own, and those of others), and historians rely on this type of ephemera to characterize an era. We are beginning our series of studies with Twitter because tweets are an important point in the space of social media, and because people have already shown signs that they want to save them (see, for example, utilities like Twapper Keeper or BackupMyTweets).

In this paper, we present the results of a survey to probe the attitudes of social media-savvy people about who can save, share, publish, and remove tweets from a Twitter feed. In so doing, we explore issues associated with ownership, authoring and reuse, and privacy. The paper is organized as follows: At the outset, we summarize related work and present the method used to construct and disseminate the survey. Then we describe the respondents and their responses to open-ended questions about their Internet use, their concerns about sharing information, and the distinctions they make between publishing material and sharing it within a social network. We then discuss our findings, identifying trends in the data and areas of controversy and consensus. We conclude by discussing implications of these findings and by outlining future work.

RELATED WORK

There are three bodies of related work which form scaffolding for this study: (1) a growing literature about Mechanical Turk and techniques for using it to recruit and screen study participants and perform studies; (2) descriptive studies of Twitter and other social media in use; and (3) studies of the social side of personal archiving, and media reuse and remix.

In recent years, researchers have found that Mechanical Turk provides a convenient pool of respondents for studies, provided that emerging best practices are taken to heart. Kittur et al. recommend that verifiable questions should be included in surveys; that good-faith efforts should pay off; and that researchers should give themselves multiple ways to detect fraudulent responses [14]. In our study description, we outline how we have designed a survey that adheres to these suggestions; in designing our payment scheme and recruiting and screening techniques, we have also taken care to follow the pragmatic suggestions offered by Jakobsson [12] and Downs et al. [4] for screening participants and avoiding cheaters.

Unlike other studies intended to characterize microblogging as a phenomenon (primarily exemplified by Twitter) [3,7,8,13,15], its use in contexts such as education [6], conferences [20], emergencies [10], and the workplace [22], we are more interested in the long-term status and value of tweets as a representative (if limited) type of social media. Previous descriptive studies have found that people tweet for a variety of reasons, some of which do produce ephemera (such as 'daily chatter'), and others of which produce content of more enduring value (such as shared information) [13]. Studies also show that people who tweet may go to some effort to ensure their microblogging posts are not ephemeral (by using them to pass information [7] or opinions [8], or even by the way in which they capture crisis-related events [10]); by adhering to retweeting conventions, they ensure that authorship is respected [3].

Personal archiving is a social phenomenon that is becoming increasingly rooted in social media. Not only are some members of families or social groups more likely to curate the group's digital assets [19], the stewardship skills required to care for these assets are often held by different people than those with the necessary curatorial skills and motivations [18]. Considerable attention has also been devoted to the emergence of a broader digital content reuse and remix culture [16]. Both personal archiving and social media reuse are largely dependent on how the rights we discuss in this paper are resolved; the main contribution of this work is a survey of digital content ownership attitudes.

STUDY DESCRIPTION

We conducted a survey to determine participants' attitudes about saving, reusing, and removing Twitter tweets; the survey also gathered self-reported information about the participants' Internet use and publication practices.

We used Mechanical Turk to recruit and screen participants for our study, as well as to implement the survey itself. The survey took the form of a series of brief scenarios that embodied specific Twitter practices, statements about these scenarios for the participants to assess according to a seven point Likert scale, and a limited number of open-ended questions to ensure that the participants were reading carefully. Open-ended questions, coupled with multiple-choice responses, also enabled respondents to describe their own Internet use and publication practices.

There is a growing body of literature that describes best practices for performing studies using Mechanical Turk [4,12,14]. We carefully followed the advice offered in these papers, and used them to set our own expectations about what to include in the survey, and how to organize it effectively. Our strategy for using Mechanical Turk was generally successful.

We couched our screener and questionnaire as a Mechanical Turk Human Intelligence Task (HIT). The HIT was offered to Turkers from the United States who have proven themselves reliable in past work (i.e. 95% or greater HIT approval rate); the HIT was framed with a strongly worded statement that we were seeking participants who were fluent in English and were experienced Twitter users.

We double-checked these characteristics in the screener; data from respondents who appeared to have only a passing familiarity with Twitter were discarded. Respondents were paid at established rates for Mechanical Turk, which worked out to 50 cents/questionnaire; respondents were paid even if their data was eventually discarded.

The Mechanical Turk HIT consisted of 34 questions. Eight of the questions characterized the respondent; six questions were intended to measure the respondent's familiarity with Twitter; three questions tested the respondent's reading comprehension (a tactic other researchers have suggested to ensure that respondents are reading the survey before they fill in responses [14,12]); 16 questions were statements of belief for respondents to assess on a 7-point Likert scale; and the final question probed the respondent's willingness to participate in future surveys and interviews. Additional data (e.g. the respondent's work time) was collected by the Mechanical Turk infrastructure.

The Six Scenarios

The belief portion of the HIT consisted of six short scenarios or situations involving Twitter. In each, users either save, retweet, or reuse their favorite tweets in other venues. The scenarios are followed by statements of belief for respondents to assess. Each scenario is specific, and includes the actual tweets. By posing concrete situations, we hoped to put all respondents on a more even footing (so they envision similar situations) and to draw on their real experiences. We also hoped that specific content and details would help to mitigate gross inconsistencies between attitudes and behavior, such as those discussed at length in the privacy literature [1]; nonetheless, we are careful to note that the data we collected reflects attitudes and beliefs rather than behavior.

In Scenario 1, a Twitter user, Dave, collects humorous tweets posted by different people, including himself; in Scenario 2, Dave collects tweets that represent a conversation he is having with two other Twitter users. In Scenario 3, Dave encounters an offensive tweet about himself. In Scenario 4, the respondent has received Dave's list of funny tweets via email, and we investigate what he or she can do with the list subsequently. Scenario 5 covers the situation in which the respondent has removed his or her own tweet from the Twitter feed, but while it was still public, Dave has collected it, and is now posting it. Finally, in Scenario 6, the Library of Congress acquires the entire Twitter archive, and provides access to it under three different conditions.

Figure 1 shows a portion of Scenario 1, followed by an excerpt of the belief statements that followed the scenario.

Four Rights of Ownership

To design the scenarios and associated statements of belief, we assumed that different situations might elicit different responses. The data will indicate which situations can be distinguished from one another. It will also indicate trends in responses to individual questions: that is, do respondents

---

Dave Sanders is one of your followers on Twitter. He tweets under the name @NewJerseyDave and you tweet as @SecretSquirrel. Dave keeps a list of his favorite tweets. This list includes some of his own tweets, some of your tweets, and some tweets posted by other witty folks. Part of the list looks like this:

SecretSquirrel A city is only really home when you stop being mystified by its public transport and instead are just constantly angry at it.

RedRabbit My personality results came back. They're negative.

NewJerseyDave I'll go to the co-dependency workshop if you come with me.

SecretSquirrel I might like the iPad better if it were red & furry.

rrmutt I know how you feel, hon: "I'd stage-dive but I'm far too elderly"—Courtney Love

NewJerseyDave Maybe the person in the car ahead of you isn't driving erratically because he's texting; maybe he's just knitting.

Dave should only be able to store his own tweets on his hard drive.
Dave has the right to post his list of funny tweets to his Facebook wall.
Dave has the right to publish his list of funny tweets in a blog post.

Figure 1. A portion of Scenario 1 exploring views about saving, sharing, and republishing a collection.

tend to agree (or disagree) with the statement (is there a normal distribution of responses), or are they divided into communities at opposition (is there a bimodal distribution)?

The following definitions of the four data ownership terms were provided to participants before they began the questionnaire to help ensure consistent interpretation:

Save – to store the content on your own storage media. For example, you might ***save*** a photo to your local hard drive or burn it to a CD.

Share – to make the content available to a limited set of friends or family members by using email or social media websites. For example, you might ***share*** a photo with your friends on Facebook.

Publish – to make the content available to the public by uploading it to a website like Flickr, Blogspot, or YouTube. For example, you might ***publish*** a story to your blog or ***publish*** a video to YouTube.

Remove – to 'unpublish' content; to delete content from a public website. For example, you might ***remove*** a photo from Flickr if you don't want everyone to see it.

Because we wanted the respondents to react to each scenario without perceiving a pattern and using it as a shortcut to fill in answers, we made concrete statements about the situation, and varied the ownership rights we tested. Again, refer to Figure 1 for examples.

Deploying the HIT

The HIT was deployed for two weeks; the bulk of the data was collected during the first week. We screened the responses based on a three point disqualification test, where a respondent must have fewer than three of the following disqualifying characteristics:

- respondent reported fewer than 5 followers on Twitter;
- respondent reported following fewer than 5 people;
- respondent reported posting fewer than 20 tweets;
- respondent spent less than 5 minutes (300 seconds) completing the survey;
- respondent answered an open-ended Twitter "how-to" question incorrectly;
- respondent answered either of two reading comprehension questions incorrectly (a third ambiguous comprehension question was discarded).

Our primary concern was that the respondents were sufficiently familiar with Twitter and its conventions to understand the scenarios and to respond to the hypothetical situations in ways that reflected their values. We were generally conservative about data quality and arrived at the formal criteria for disqualifying respondents only after we took a close look at the data we had collected to identify possible ways in which respondents had gamed the system.

At the end of two weeks, the HIT had been completed by 190 respondents; 173 passed the secondary screening criteria. Respondents took on average 8.7 minutes (8 minutes, 42 seconds) to complete the survey, which was sufficient to read and understand the scenarios and statements, and to complete the open-ended questions.

One measure of the HIT's success is that the answers to the open-ended questions were surprisingly rich; it was clear that respondents took the survey seriously, and offered us a real window onto their attitudes and their practices.

CHARACTERIZING THE RESPONDENTS

Eight initial questions were aimed at characterizing the study participants. From the screening, we knew the participants were familiar with Twitter, and because they were recruited from the Mechanical Turk worker population, we also knew that by some measure, they were Internet-savvy. Although there are characterizations of the Turk workforce [11], we thought it was important to describe the subpopulation of respondents.

Respondents were largely in their twenties (64%) and thirties (17%); 61% are female, 39%, male. This is unsurprising, given the characteristics of US Mechanical Turk workers. Table 1 breaks down the age and gender distribution of the participants in our study.

year born	before 1950	1950-1959	1960-1969	1970-1979	1980-1989	after 1990	declined to state	total
female	1	4	3	16	69	11	1	105
male	0	2	4	13	41	8	0	68
total	1	6	7	29	110	19	1	173

Table 1. Participant age and gender

The majority of the participants are college-educated—94 (54%) have a college degree—and almost all of them (152, or 88%) have at least some college education, including those who are still students. Unfortunately, we realized after the fact that we did not distinguish college dropouts from current students, but we assume that a significant portion of those who characterized themselves as having some college (58, or 34%) were probably current students. Twenty participants (12%) reported having a graduate degree.

Participants reported their familiarity with social media applications in multiple ways. They specified the number of years of Internet experience they had; selected which of a list of nine types of social applications that they use; and described in their own words how they use the Internet, what they spend the most time doing, and types of social media they create themselves.

Table 2 shows the distribution of application popularity; the ninth application was Twitter, which all qualified respondents reported using. Email and social networking are respondents' most commonly used applications.

Social media activity	# of respondents
Email	171
Social networking (e.g. Facebook)	161
Shopping (e.g. Amazon)	131
IM/chat	122
Online video sharing (e.g. YouTube)	110
Online photo sharing (e.g. Flickr)	84
Video conferencing (e.g. Skype)	60
Massively multiplayer online games (e.g. WoW)	37

Table 2. Popularity of applications among respondents.

Participants engage in a broad range of online activities; the majority have used the Internet for a long time, sometimes for a significant portion of their lives, and most engage is four or more social media activities (see Figure 2).

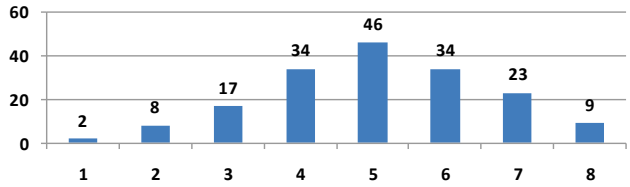

Figure 2. Number of social media activities (X axis) by number of respondents (Y axis).

Although the respondents were by no means homogenous, they seem to represent a type of Internet user—educated, fairly young, with a spectrum of attitudes about topics like privacy, and a range of Internet interests, from buying and selling in online marketplaces (e.g. Etsy and Craigslist) to playing multiplayer games to tracking news to participating in social networks (e.g. Facebook).

Internet Activities

Open-ended questions included a broad question about how respondents used the Internet, and what they spent the most time doing. Many of the respondents offered fairly detailed information about their extensive and varied Internet use; most cite communication-related activities as a principal preoccupation. For example, R77 said,

> "I use the Internet for things like communicating with friends and family (some nearby and some very far away) via email, IM, and social networking, I do research and shopping, I use it to streamline managing my finances, I follow hobby-related groups, and use it to pass the time when I'm bored. I'd say that I spend the most time communicating with friends and family on the Internet."

Respondents were also very specific about pursuits enabled by the Internet, often focusing on what they use the Internet for (growing as a Muslim [R159] or day-trading [R152]) rather than on applications they use or generic activities (communicating, browsing).

Unsurprisingly (since they are workers on Mechanical Turk), many respondents also report use related to some type of resourceful self-employment. For example, [R153] wrote, "As a web designer, much of my time spent online is work-related, either directly or indirectly (research)" and [R109] wrote, "I play the game SecondLife and use SL to make videos for my online business." Other respondents reported working as news aggregators (e.g. [R48]) or doing research in support of a non-Internet business, e.g. as a personal trainer [R75].

Publishing and Republishing

We asked respondents what types of information they published or shared on the Internet, with the thought that a broad framing would elicit informative responses. Although we had already brought up different types of applications, respondents went well beyond this limited range. In addition to saying that they shared photos of friends, and pictures they found, they also reported sharing things like "scans of my artwork" [R103], and republishing "pornography and videogames" [R167] and "information in Second Life" [R90].

Republishing arose in several contexts, sometimes as an offshoot of sharing articles and news stories, and in other cases as intentional piracy (e.g., when the respondent reported reposting porn and videogames).

Respondents reported creating social media, sometimes as a way of sharing their experiences with physical media ("I share projects I'm working on and helpful tips in the knitting community" [R132]; "I have a band, so I share songs pretty often" [R37]; "I write a lot of fanfiction and I use photoshop to create fanart" [R25]; "I write some game reviews and walkthroughs" [R64]).

Responses to this question included a number of new, evolving social media genres, including videogame walkthroughs, personal profiles, status updates, fan fiction and fan art, and links and blogs. The social marketplace of opinions, reviews, and Etsy and eBay postings is also well-represented in respondents' answers.

This diversity of use and sharing practices implies that respondents are conversant in social media, and are likely to have thought about the issues raised by the scenarios.

Sharing Personal Information

In both open-ended questions that preceded the scenarios, respondents volunteered concerns about sharing personal information via social media. These responses were not elicited by asking about privacy specifically; rather they arose organically in the course of respondents' reports of how they used the Internet or what they shared or published. While we cannot be certain that respondents did not try to ascertain our interests in administering this survey, there was nothing in the phrasing of the instructions that indicated that we were concerned with privacy.

From respondents' answers, it was evident that some divided social media into public and private information:

> "I try to publish/share information that would not offend my personal sense of privacy. While most stuff in my life is fair game (What I ate for dinner, what the score at the baseball was, etc.) some is not, such as how much I spent to attend the baseball game or on dinner, any information that may put someone (including myself) into legal, financial, or physical jeopardy, etc." [R77]

There was also a sense that some venues (e.g. Facebook) were more private than others (e.g. Flickr and Twitter):

> "I publish pictures and have a blog, but I don't share anything too personal. My facebook is very private only my close friends, but my flickr, twitter, and blog are public." [R158]

Other respondents expressed a sense of uncertainty and a fairly conservative outlook about life online:

> "I don't usually publish too much information about myself because I don't know who is going to look at it." [R39]

> "[I publish] As little information as possible for me to continue using the various sites that I am interested in." [R7]

> "I write articles on internet culture and technology. I tend not to write too much about my personal life." [R1]

On the other hand, some respondents indicated that they led a lively online life, that they had little concern for who encountered their personal postings:

> "I publish and share everything! Bad days, good days, what I make for dinner, when I go to bed I post goodnigts [sic], etc." [R41]

Respondents also indicated that they curated the information that they published online according to who they felt would see it; once they were publishing to a local audience, they felt they had to be meticulous. As the quotes above indicate, some respondents kept personal information to a minimum, providing only what was needed to meet the expectations of their audience and the requirements of the service. What they considered private varied, however:

> "I don't usually publish too much information such as location (city wise) and just recently I started privatizing my email address." [R5]

"Very highly curated information about my personal life; reviews of outside things, etc." [R98]

"Depends entirely on the trustworthiness of the person or site as demonstrated by time and consistently. Some know my name, address and phone number. Some only know my penname and e-mail." [R110]

FINDINGS AND DATA TRENDS

One of the aims of this survey was to tease apart closely related concepts. First, we investigated the nature of social media ownership. Who can archive a shared artifact? Ownership may be complicated by collaboration, by remix practices, and by reuse.

Next we probed the difference between publication and sharing: if a person has 500 Facebook friends, and their public blog is only read by 50 people, is there a difference between sharing and publication? And if there is a difference, what implications does it have? Furthermore, does it influence how people feel about institutional archiving of lightweight social media? If something has been published, people may have a reasonable expectation that it may be archived. Yet if it is just shared, people may still feel that it is part of a private realm, over which they may exercise the ultimate control.

Although there has been significant litigation centering on copyright infringement, we wondered whether social notions of reuse were deviating from codified laws and legal precedent. If a person collects social media—in this case tweets—does he or she have any authorial claim on the aggregate material? And what of material that was authored collaboratively, as a Twitter conversation would be—how far do reuse rights extend?

Finally, we examined how shared artifacts are unshared, and how published items are unpublished. Social media appears to deviate from the print/physical realm, where copies are more difficult to make, and media is thus less fluid and potentially volatile [17]. Teenagers are told to mind what they share on Facebook—that once something is shared, it cannot be unshared, or that a video posted to YouTube will 'live on forever on the Internet.' How do people feel about ownership and removal of social media?

We use five concepts to organize our findings:

- Social media ownership and control;
- Publication versus sharing;
- Reuse and republication;
- Removing social media;
- Institutional ownership and individual data rights.

We discuss each concept in turn. Although they are inexorably intertwined, we do our best to pull them apart, and discuss them from different perspectives.

Table 3 presents an overview of the respondents' reactions to statements about the scenarios. In the table, brief squibs that summarize the scenarios' basic premises are shown in bold. They are followed by statements that the respondents were asked to score on a seven point Likert scale. Mean and mode are shown for each statement; graphs later in this section present the full distributions.

Table 3 helps highlight statements that met with bimodal distributions of responses, and statements that are viewed in consistently favorable or unfavorable ways. For example, responses to statements about peoples' rights to store tweets, one's own or someone else's, were likely to be highly favorable (i.e. the mode value of these responses was 7, the most positive possible response). The distribution only became bimodal when we attempted to impose limits on which tweets people were allowed to store. Similarly, respondents displayed unfavorable responses to the right to delete tweets *about* oneself (see, e.g., Q12; Twitter allows users to delete tweets written *by* oneself). The Library of Congress scenario elicited bimodal responses, in particular, if the institution provides open access to the collection of tweets to the general public today (i.e. Q21). Attitudes about institutional archives appear to be very much divided; time delays (50 years in the future) and access limitations (access is granted to researchers only) ease respondents' discomfort with full public access to archived tweets.

Ownership and Control

Ownership and control over social media is an important foundation for designing many types of services. We are particularly interested in personal archiving—what will the long-term fate of social media be? Although social media is usually posted by a single author (or by a single identity), it is often mutually owned, and certainly it is often about a group or an event, or may refer to or depend on something else on the Web. The person who posts an item may not be the subject of the item, nor may the same person curate or maintain the media over time (e.g. see [19]). Who owns social media, and who controls it?

In this study, we are talking about something on the simple end of the spectrum, a tweet, a 140 character text string. Yet, as the scenarios posed in the survey illustrate, even something as simple as a tweet introduces considerable complexity. A tweet might be intensely personal; it might be plagiarized; it might be about someone else; it might refer to a social event; it might have a hypertext link that connects it to another, richer, piece of social media or external resource. In other words, it is difficult to make definitive statements about who owns and controls even lightweight social media.

The strongest positive reaction the survey elicited had to do with the storage and archiving of one's own tweets. Respondents felt strongly that they should be able to store their own Twitter conversations, especially the material that contextualizes what they have written (Q8). Although the reaction is not as strong, respondents seem to feel comfortable storing the tweets from their feed, regardless of who wrote them (Q3). (The tweets in the scenarios do retain attribution.)

Statement	*Mean*	*Mode*
Dave is one of your followers on Twitter. He keeps a list of his favorite tweets. This list includes some of his own tweets, some of your tweets, and some tweets posted by other witty people.		
Q3. Dave should have the right to store a copy of all of these tweets on his hard drive.	**5.36**	**7**
Q4. Dave should only be able to store a copy of his own tweets on his hard drive.	**4.09**	**7**
Q5. Dave hopes he can entice his friend Charlie to join Twitter and become one of his followers. Dave should have the right to post his list of funny tweets to his Facebook wall.	**5.17**	**6**
Q6. Dave should have the right to publish the list of funny tweets in a blog post.	**5.36**	**6**
Suppose Dave's list just contains tweets that are his own tweets and responses to them.		
Q8. Dave should have the right to store a copy of all of these twitter chats on his hard drive; after all, they're all either by him or about him.	**5.75**	**7**
Q9. Dave should only have the right to store a copy of his own tweets on his hard drive, not the responses written by other people, even if it makes it hard to understand the conversation.	**3.38**	**2**
Q10. Dave should have the right to post the entire conversation to his Facebook wall so his friends can see it.	**5.01**	**6**
Q11. Dave should have the right to publish this list of tweets—tweets that he wrote and responses to tweets he wrote—in a blog post.	**5.27**	**6**
Dave sees a tweet that claims he's gained 50 pounds; the tweet's author does not realize Dave is on Twitter and will see the tweet.		
Q12. Dave should have the right to delete the offensive tweet about him.	**3.71**	**1**
Suppose Dave emails his list of favorite tweets to you. They're hilarious. What can you do with the list?		
Q13. You should have the right to email the tweets to a friend without leaving Dave's email header on the forwarding.	**4.45**	**7**
Q14. You should have the right to publish the tweets in your public blog.	**4.41**	**6**
One of your tweets that's on Dave's list embarrasses you. You had deleted it from Twitter not long after you'd posted it, but Dave caught it while it was still part of the feed. Now he posts the whole list on Facebook.		
Q15. Suppose you can't delete only part of Dave's post. You should have the right to remove Dave's whole Facebook post to get rid of your bad tweet.	**3.09**	**2**
Q16. Suppose you can edit Dave's post. You should have the right to remove your bad tweet from Dave's published list. After all, you removed it from the Twitter feed.	**4.32**	**6**
The Library of Congress is acquiring the entire Twitter archive, dating back to its origins in 2006.		
Q20. The Library of Congress can give researchers access to the archive.	**4.62**	**6**
Q21. The Library of Congress can give everyone access to the archive.	**4.10**	**1,6**
Q22. The Library of Congress can give everyone access to the archive after 50 years has passed.	**4.40**	**6**

Table 3. Overview of the sixteen statements of rights and the Likert-scale responses to them.

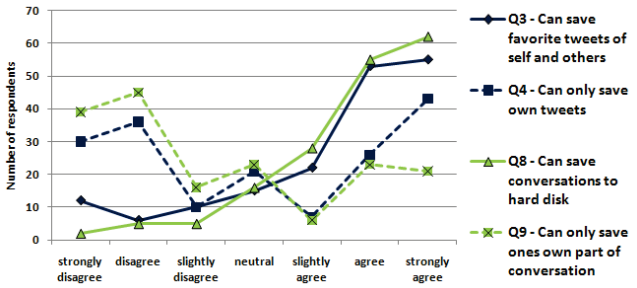

Figure 3. Views about saving favorites and conversations.

Publication versus Sharing

From its inception, social media introduced the idea that we don't just publish personal content to the open Web. Rather we share and build on existing content among our spheres of family, friends, colleagues, and acquaintances, making ever more nuanced distinctions between who sees what and what they are allowed to do with what they see [21]. Are people developing varied sensibilities about the distinctions between publication (posting social media on the open web) and sharing (controlling who sees what)?

In our Twitter scenarios, one way to investigate this question is to test statements about posting a collection of tweets to one's Facebook wall (as a form of controlled sharing), and publishing the same collection as a blog post.

Note that the collection aggregates material from different sources: the tweets were written by different authors, and collected into a list of favorites.

A surprising finding is that respondents approach sharing on Facebook with more caution than they do publishing the same content on a blog. The difference between the scoring of the statements that embody this distinction is statistically significant ($0.02<p<0.05$, Wilcoxon Signed Rank Test) although the difference is not large (see the publication statements Q6 and the sharing statement Q5 in Table 3).

From the perspective of sheer exposure (the people who can see the Facebook post are necessarily a subset of the general Web-reading public who can see the published blog), this result is counterintuitive. We explain this difference cautiously, since there is nothing in the data that pinpoints the reason for it. We speculate two effects are at the root: (1) the Facebook audience is an audience of friends, and the blog audience is potentially an audience of strangers, and people are often more careful about what they share with a captive audience of Facebook friends (who may be more or less socially obliged to read the post, and may judge the person who posted it accordingly) and the self-selected audience of blog readers (who, the author may rationalize, are free to 'vote with their feet', and who may be less judgmental about whether the author has the

right to publish the tweets) and (2) experience tells us that a blog may in fact have fewer readers than a Facebook wall post, a venue that suggests social engagement.

This result is consistent with one revealed by a second hypothetical situation involving a Twitter conversation rather than a collection of unrelated tweets (embodied by statements Q10 and Q11). Again, Facebooking the conversation was seen as an action to be approached with greater caution than simply publishing it broadly to the Web (as a blog post) ($p<.001$, Wilcoxon). These statements about sharing and publishing differed in a statistically significant way, while the pairwise comparison of the similar statements (sharing a collection vs. sharing a conversation and publishing a collection vs. publishing a conversation) were not significantly different using a Wilcoxon test ($0.05<p<0.1$ and $p>0.2$ respectively).

Thus our results raise a further question: is this a broad social media phenomenon? Where data ownership is concerned, is sharing really something to be approached in a more cautious, self-conscious way than publishing? In other words, do people actually worry more about their rights to post a photo to Facebook than they do to the open Web? This result seems worth further investigation; that a person would worry more about whether he or she actually has the right to post something in a social situation than he or she does in an abstract information environment makes sense; the effect may be even more dramatic when other types of media (such as photos) replace the lightweight (and often emotionally distant) media forms like tweets.

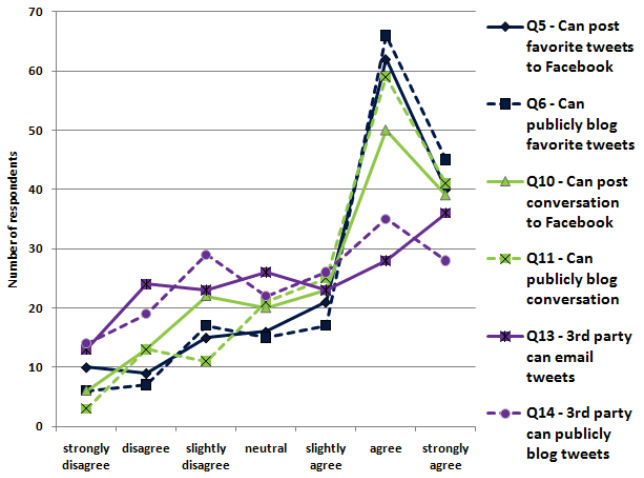

Figure 4. Views about sharing favorites and conversations.

Reuse and Remixing

Reuse and remixing represent emerging authoring practices for the Internet age [16]. Reuse and remix imply that the author is gathering media, excerpting desired portions, and putting them together in new and creative ways. There is a lingering feeling in some quarters that this practice edges on plagiarism; copyright law often reflects this sense of unease. But this feeling is by no means pervasive. We thought it was worth investigating reuse through our scenarios, and respondents brought up the issue themselves in the portion of the HIT that preceded the scenario-based survey statements and asked about respondents' practices.

For example, respondent R131 said, "Content I share is usually not my own, I just retweet, digg, or share it on facebook in hopes others will see in the piece the same thing I found." Others explained reuse in a less elaborate way; they unselfconsciously reported publishing or sharing:

- "Music, interesting or funny pictures I come across, videos, jokes" [R106]
- "sharing links/articles that I find interesting" [R159]
- "movies and videos. pornography and video games." [R167]
- "Pictures that I find in the internet." [R118]
- "I publish humor stuff on my network of web sites." [R59]
- "Information about …deals I find, or interesting articles I come across." [R139]
- "previously published information" [R129]

It is then not surprising that respondents incorporated this right in their reactions to the scenarios (Q5, Q6, Q10, Q11, Q13, Q14). All of these scenarios involve either combining one's own material with tweets written by others, or simply reusing the kind of material that notoriously floats around the Internet as 'funny email', blog posts, tweets, or in other forums. The statements that involved sharing or publishing material without attribution to the collector who reused them (Q13 and Q14) scored significantly lower ($p<.001$, Wilcoxon for all comparisons) than the tweets that were attributed to their authors (but still republished without permission in venues other than Twitter) (Q5, Q6, Q10, Q11).

Removing Content

Constructive reuse (storing, publishing, sharing) was greeted less cautiously than removal. Instead of testing removal without any particular motivation, we introduced 'aboutness' to ownership: i.e. what are the limits of what you can do with material that is not by you but rather about you? People may feel they have a certain latitude to take action on materials that are about them, media documenting an event they attended, or using them as the subject. They feel free to tag themselves or other people they know in photos, for example, or they feel it within their rights to request that a photo of them be taken down (and surely in some cases, they would have taken it down themselves had they been able) [1].

Since tweets are so lightweight, we first tested removal using a sensitive topic, a tweet reporting that Dave looked as if he had gained 50 pounds. Can Dave remove such a tweet (as posed by Q12)? Responses to straightforward removal of this sort—when the tweet was about but not by oneself—revealed it to be a controversial action.

It would have been less controversial if the tweet was authored by the respondent, since this is a capability already offered by Twitter. We complicated this scenario (removing one's own tweets) by introducing the notion of a tweet that the respondent had already deleted using Twitter's delete function, but that had been reused by a follower before the deletion occurred. Did respondents feel they should be able to take matters into their own hands and make sure the delete propagated to the places where the tweet was reused? In other words, should the author be able to remove the reused tweet from the alternate contexts in which it now appears? In one of the situations we proposed, this removal could be performed surgically without further damaging the aggregate post that the author did not create. Respondents' scores trended mildly positive (see Q16 in Table 3). However, if the removal destroyed surrounding material too, the response trended negative (see Q15 in Table 3). This difference is significant ($p<.001$, Wilcoxon).

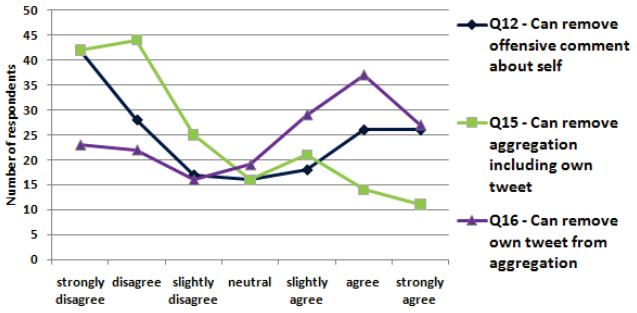

Figure 5. Views about removing tweets and aggregations.

Institutional Ownership and Individual Data Rights

Not long before we conducted this survey, the Library of Congress acquired the entire public Twitter database as part of an initiative to archive Web-based information. This acquisition was not without controversy: some Twitter users felt they had achieved a measure of privacy through obscurity, and although in principle their tweets were already available to the public, they were viewed and used in a known context, under an understood set of social mores. Yet it is part of the Library of Congress's mission to collect such ephemera, so that future scholars and historians have a window onto our time.

It seems worthwhile to investigate our respondents' attitudes toward the rights of public institutions to collect and archive social media such as tweets, so we posed the Library of Congress's actions as hypothetical, and in terms of access rather storage: Does the Library of Congress have the right to give the public access to this collection of tweets? What if the access were limited to researchers? What if public access were deferred for 50 years?

Interestingly, the most positive response was to the case that limited the access to researchers. This hypothetical limitation defined the context of use, and seemed to ease discomfort with the idea of full public access to tweets. The other two variations—full current access and full access in 50 years—elicited bimodal distributions of scores, although according to the mean scores, granting full access in 50 years and limiting access to researchers today fared similarly. The difference between Q20 (researcher access) and Q21 (full access) is significant ($p<.001$, Wilcoxon).

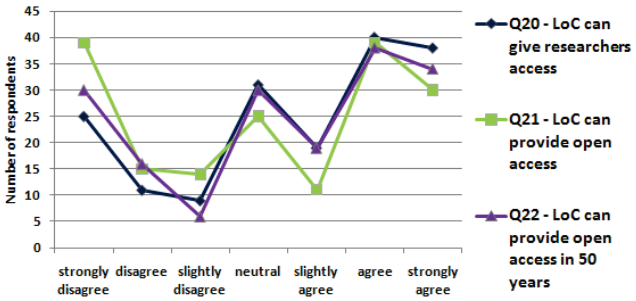

Figure 6. Views about institutional access to tweets.

CONCLUSION AND FUTURE WORK

What do these attitudes toward ownership rights tell us about Twitter content, and more generally, about social media? How should attitudes about social media ownership inform the design of collaborative services, especially services such as personal archiving? We approach this connection with a measure of caution, since we know there can be an important gulf between attitude and action, as demonstrated by privacy research (for example, [1]).

Yet we feel that attitudes are a good place to start, especially if they are coupled with investigations of existing behavior. For example, past studies tell us that people often approach the long term storage of their digital stuff with a strong tendency toward benign neglect [18]: if an archiving service were offered, where would the boundaries on stored content be? What would people be able to do with the contents of such a personal archive?

Much of the personal material that is on the Web is jointly owned—it is created collaboratively, records social settings and events, and is curated on the social Web. In other words, personal archiving is no longer a matter of creating a permanent copy of material from a personal computer's hard drive. In Twitter, a tweet may be written by one author, may refer to someone else (either about them or their content), and may implicate others as well (for example, as a retweet). Similarly, a photo on a photo sharing service may be taken by one photographer, be of several people in the foreground, and others visible in the background; it may further be described by metadata contributed by other people. In general, social media is a social endeavor, and hence it worthwhile to consider how this joint ownership affects its long-term fate.

There is some consistency to actions respondents feel they can take with lightweight social media like tweets. Certainly there is a sense that one can store almost any Twitter content that one encounters. More caution is exercised with publishing and republishing tweets, and more still with sharing the tweets among friends and

associates. Finally, the most caution is exercised with the right of removal: deleting tweets seems to be seen as a potentially antisocial act. Even when it is not destructive, some respondents feel that they should exercise all due caution. It is important to know whether these findings extend to other types of social media, to begin to explore the ethos behind these beliefs, and to investigate the relationship between these attitudes and lived practice.

In this study, we found that the respondents' relationship to the material influences what they feel they can do with it. If they are cast as the author, they are the least cautious. They are slight more cautious if the tweets are cast as public (where there is no strong authorial claim laid to the material), which is in turn followed by tweets that reflect shared ownership. The most caution is associated with participatory content (e.g. Twitter conversations). Again, does this ordering hold for other types of social media?

Some ideas we posed in this survey seem to provoke disagreement or controversy. We uncovered three: (1) Respondents disagreed about limits that might be imposed on what you can store out of your twitter stream based on the individual tweet's role and provenance; (2) Respondents disagreed about whether public or governmental institutions should be able to provide universal access to public tweets; and (3) Respondents disagreed about whether individuals should be granted the power to remove tweets about themselves (tweets they did not write) that are potentially unpleasant or libelous.

Our future work will determine whether these belief patterns will hold for other social media types, particularly those that require more authorial effort or have greater emotional currency (e.g. family photos or videos). Further investigation will resolve whether these patterns of action (store, publish, share, remove) or authorial relationship (author, public content, shared content, participatory content) are meaningful in other social media contexts, and will identify potential gaps between attitudes and behavior.

REFERENCES

1. Acquisti, A. and Grossklags, J. Privacy Attitudes and Privacy Behavior, in J. Camp and S. Lewis (Eds.) *The Economics of Information Security*, Kluwer, Boston, pp. 165-178.
2. Besmer, A. and Lipford, H.R. Moving Beyond Untagging: Photo Privacy in a Tagged World. *Proc. CHI'10*, 1563-1572.
3. boyd, d., Colder, S., & Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *Proc. of HICSS*, 2010.
4. Downs, J., Holbrook, M., Sheng, S., & Cranor, L. Are your participants gaming the system?: Screening Mechanical Turk workers. *Proc. of CHI'10*. 2399-2402.
5. Dourish, P. and Bly, S. Portholes: supporting awareness in a distributed work group. *Proc. CHI '92*. 541-547.
6. Ebner, M. & Schiefner, M. Microblogging - more than fun? *Proc. of IADIS Mobile Learning Conference*, 2008, 155-159.
7. Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., & Kellerer, W. Outtweeting the Twitterers - Predicting Information Cascades in Microblogs. *Proc. of Workshop on Online Social Networks*, 2010.
8. Honeycutt, C., & Herring, S. Beyond microblogging: Conversation and collaboration via Twitter. *Proc. of the HICSS 2009*. IEEE Press.
9. Huberman, B.A., Romero, D.M. and Wu, F. Social Networks that Matter: Twitter under the Microscope. First Monday, 14, 1, January, 2009.
10. Hughes, A.L. & Palen, L. Twitter adoption and use in mass convergence and emergency events. *Int. Journal of Emergency Management*. 6 (3/4) 2009, 248-260.
11. Ipeirotis, P. The New Demographics of Mechanical Turk. http://behind-the-enemy-lines.blogspot.com /2010/03/new-demographics-of-mechanical-turk.html
12. Jakobsson, M. (2009) Experimenting on Mechanical Turk: 5 How Tos. *ITWorld*, September 3, 2009.
13. Java, A., Song, X., Finin, T. & Tseng, B. Why we twitter: Understanding microblogging usage and communities. *Proc. ACM SIGKDD*, 2007.
14. Kittur, A., Chi, E., Suh, B. Crowdsourcing User Studies with Mechanical Turk. *Proc. CHI'08*. 453-456.
15. Krishnamurthy, B., Gill, P., & Arlitt, M. A Few Chirps about Twitter. *Proc. Workshop on Online Social Networks*, 2008, 19-24.
16. Lessig, L. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, Penguin, New York, 2008.
17. Levy, D. "Fixed or Fluid: Document Stability and New Media." *Proc. ECHT'94*, 1994, 24-31.
18. Marshall, C.C. Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field. *DLib*, 14, 3/4
19. Odom, W., Harper, R., Sellen, A., Kirk, D., Banks, R. Passing on & Putting to Rest: Understanding Bereavement in the Context of Interactive Technologies, *Proc. CHI'10,* pp 1831-1840.
20. Reinhardt, W., Ebner, M., Beham, G., & Costa, C. How People are using Twitter during Conferences. *Proc. of EduMedia*, 2009, 145–156.
21. Stutzman, F. and Kramer-Duffield, J. Friends Only: Examining a Privacy-Enhancing Behavior in Facebook. *Proc. of CHI'10*, 1553-1562.
22. Zhao, D. & Rosson, M.B. How and why people Twitter: the role that micro-blogging plays in informal communication at work. *Proc. of GROUP'09*, 243-252.