Alex Carson
Department of Anthropology
University of Maryland College Park

Informing Consent

A study on privacy and consent in internet research.

Question: Should researchers feel free to draw from "public" information found on the internet?

Background

As the internet has become an increasingly prevalent mode of communication, particularly in industrialized nations, it has also become an increasingly common means through which academic research is carried out. With less than twenty years of what would truly be considered broad accessibility to the public – and even less time in which phenomena like social networking have driven individuals to publish information regarding themselves in potentially public spaces – research on the internet is evolving as fast as its medium, and research ethics in turn have attempted to keep pace with this dizzying rate of change.

One of the most prevalent and important aspects of ethics in any research is in the treatment of private and public information, and since the rise of academic research utilizing the internet and information found on it there has been an uneasy status quo on the subject. Early academic writings on internet privacy espoused a simple standard in which information communicated in a space open to the public could be considered public, whereas information exchanged strictly between two individuals should be considered private and subject to academic protocol relating to the acquisition of informed consent per Institutional Review Board guidelines (Source).

While this definition can be and remains consistently used in the course of academic research utilizing the internet, current events have arguably called into question the manner in which researchers should handle all data and, specifically, "public" data. In these situations, data which was made "public" was utilized or could have been utilized in a manner which had or would cause harm to the party to which the information initially belonged. The damages which were or potentially could have been caused through this data include damage to the dignity of the individual, damage to the user's financial well-being, or even damage to the user's safety. These case studies are provided in the interest of demonstrating the current problems with conceptions of "private" and "public" information and how this information is handled by researchers.

Case Study #1: The AOL Search Data Leak

Overview: On August 4th, 2006, AOL Research released a compressed file documenting millions of search terms entered into AOL by several hundred thousand of its users obtained over three months. While the information associated with these searches, such as the IP address, was anonymized and the users given random numbers rather than names or other indications of identity, individuals around the world who had unlimited access to this data were able to find several means – searches of locations, searches of one's name or other identifying information, etc. – to identify (source) or at least approximate the identity of those individuals whose data was taken. In another questionable use of this data, a broadway play has been commissioned about one of the users due to his or her strange search entries (source).

Alex Carson
Department of Anthropology
University of Maryland College Park

Ethical Questions: The primary ethical problems regarding the AOL Search Data Leak are rather apparent and have entered the academic discourse since the incident occurred (source). The value of research versus the damage it could cause is not the purpose of this case study, though. Rather, this case study is designed to illustrate that information which is "publicly" available may not be so by the user's consent or even legally, as it has been alleged that the release of the data was a violation of AOL's own terms of use (source). Thus, the traditional notion of information available without special access being public should be called into question by events such as this and the following case study.

Case Study #2: Google Buzz

Overview: On February 9th, 2010, Google announced the opening of Google Buzz (source), Google's first foray into social networking built out of the infrastructure for Google's popular e-mail service. In implementing the service, Google allowed others on a user's "frequent contacts" list to freely access certain information about them even if the user in question had not signed up for Google Buzz. This issue was brought up in an article politely entitled "F*ck You, Google" (source) in which a writer reported that in Google Buzz's implementation, which she had not been signed up for, her personal information had been made available to her abusive ex-husband and no recourse to block him had been provided by Google in Buzz's implementation.

Ethical Questions: In this case study, the user's information was released without her knowledge or consent due to the "opt-out" nature of Google Buzz, in which Google users were automatically signed up to Google Buzz and rather had to "un-join", which has been noted as a very difficult process (source). This was done as a part of a decision by the Google Corporation, which sought to compete with social networking websites and inadvertently released this information as a part this policy and in order to promote their social networking service to members. It should also be noted that in the last case, where information was hypothetically anonymized, the information made public by Google was tied to any data the user had provided to Google and, in this case, provided the potential for the user – an abuse victim – to suffer emotional or even physical harm. While Google corrected the mistake over the coming weeks and the Google Buzz service ended in January of 2012 (source), we refer to Case Study 1 and the functional permanence of information on the internet due to the ability of internet users to download information. Indeed, the information contained on Google Buzz remains accessible through Google's other services.

Case Study #3: Job Interviews and Facebook

Overview: In the past several years, increasing attention has been paid to the practice of employers checking potential employees' names on Google and other search engines. One article tells of a student named Felicia (source) who, in the process of seeking employment, typed her name into a search engine and came up with controversial, embarrassing, and ultimately potentially harmful information about herself from her un-secured social media use. Even with

the information made private, though, Felicia found that it was still available on other websites which had saved the information on their own and was still accessible.

Ethical Questions: This case study is last before it has the most relevance to the data gathered and mentioned in the second section of this overall study. In this case, we see that information which was made public by a user could cause harm them in a tangible sense through the story of Felicia and her "net trail", as the article refers to it as. Felicia's response to the knowledge that her information was public and easily accessible calls into question the notion that "public" data is "public" by intent. It is also worthy of note that the permanence of information posted on the internet has appeared now in all three case studies and will be addressed in the remainder of this overall case study.

Facebook's Timeline

In light of the issues raised in the above case studies regarding online privacy, it is reasonable to question current academic standards relating to the collection of "public" information and to what degree internet users intend for this information to be publicly available. While these cases do provide room for doubt, however, they do not provide a framework for addressing these doubts and establishing an ethical framework for internet research which protects a participant's privacy and addresses safety issues related to privacy in the most thorough manner possible.

In order to contribute to this framework, an anthropological study was conducted over the summer of 2012 to explore the issue of online privacy as it related to Facebook's Timeline feature. Facebook's Timeline feature was chosen because of the controversy surrounding it. Upon the release of the feature, many users and outside observers complained that it dredged up and made public information which was long thought by many users to be gone and forgotten , and as such "private" (source). While there were other complaints regarding security problems that came along with Timeline, however, Timeline itself did not make available any previously private information, but rather made access to past information easier. This lends Timeline to being both a contemporary and contentious issue, but due to the fact of no new information truly being available, one which the investigation of will pose minimal risk to users.

The research was conducted via ten semi-structured interviews with friends and associates of the researcher carried out in late July and early August. The interviewees consisted of 5 males and 5 females between the ages of 20 and 56. 7 of the respondents identified as Caucasian or white, while the other 3 identified as being Asian. Of the 10 interviews conducted, 7 were conducted by e-mail, 2 by in-person interviews, and 1 by Facebook's messenger program. The pseudonyms, as well as basic demographic information for respondents, will be provided below.

An analysis of these interviews, including several excerpts from them and critiques of the methods applied in carrying out this research, will be summarized in the following section. To protect the privacy of the users, all will be referred to be aliases and no personal information tying their aliases to their identities shall be disclosed. The analysis will be presented on a

question-by-question basis, with trends in the answers being discussed with each answer and in summary at the end of the analysis.

Basic Information

Amy C: 24, Female, Caucasian
Johnny G: 56, Male, Caucasian
Juan C: 20, Male, Caucasian
Leonard B: 20, Male, Caucasian
Liam N: 49, Male, Caucasian
Maya A: 22, Female, Caucasian
Michael A: 24, Male, Asian
Sally E: 25, Female, Asian
Susan D: 23, Female, Caucasian
Lynn V: 36, Female, Asian

Critiques: While the purpose of this research was not to gain an exact, representative sample of Facebook users, I did seek out users who I knew to represent a broad section of Facebook users. In retrospect I should have acquired more background information to break these users down by than these factors, as I know things about them that I unfortunately did not ask about and should have, but with regards to age and sex I believe I acquired a very broad sample. One aspect where I believe I could have done better is in race, where representing an African-American or a Hispanic individual may have contributed valuable insights into the project.

As for the participants sampled, I believe there were advantages and disadvantages to drawing primarily from friends and professional and recreational associates. Of the advantages, I believe the greatest were the ease of finding 10 participants and, I hope, their willingness to speak frankly and honestly to me based on our past associations. The disadvantages become apparent, however, in the demographic shortcomings of the study due to my own, limited social circle. A bit more diversity would have been ideal, but I believe that the study ultimately remains valid. Another disadvantage may be the flip side of the participants knowing me personally, as I had concerns they might feel the desire to "perform" for me or give me answers I desired. This concern was somewhat soothed, however, as several of the respondents who knew me best gave me some of the answers most contradictory to my personal views on the subject.

<u>Question 1: How would you describe your level of proficiency with the internet? With electronic devices such as computers or phones used to access it?</u>

Most participants in the interview consider themselves to be "average" or "proficient" users, with 6 describing themselves as "very" or "pretty" proficient (Amy C, Liam N, Sally E, Michael A, Juan C, and Susan D). Susan D identified herself, on a 1-10 scale, as being an 8.5 (with 0 being not proficient and 10 being very proficient). Of the others, one professed to being an "average" user (Johnny G) while the other three expressed the general sentiment that they had a basic

proficiency (Leonard B, Maya A, and Lynn V), but did not necessarily know the underlying mechanics of the technology they were using. While I could not find any patterns on my own relating to the demographic information, the interview with Juan C shed some light, as he explained that his higher education necessitated proficiency with technology in order to pass his courses. This may be a common element between the types of responses I received, and further investigation may be warranted into determining if there is a recurring link between education and proficiency with using online media such as Facebook. This data will be valuable in what I believe to be the important task of determining if there is a correlation between one's familiarity with technology and their concerns over security.

Critiques: While scales will always carry a degree of subjectivity, I think building a scale like that mentioned by Susan D into the study would have given me a better way to relate to data, as at the end of the day "pretty proficient" or "very proficient" only mean something in the context of the rest of the answer. While this is, at the end of the day, qualitative data, a more basic standard might have been helpful in analysis. Also, I feel that I may have gained something from requesting the user's education level in the demographic information to better discuss trends in an issue that may have a lot to do with one's level of institutional education.

Question 2: How do you use Facebook? What purpose do you utilize the website for?

This question had two purposes. The first was to assess the diversity of the participants and determine, as the question would imply, how they choose to utilize Facebook. The second, which in fact only became apparent to me as I carried out the interviews, ended up being to determine what sort of risks were posed to the participants in making this information public. While two of the participants used Facebook for both personal and business uses (Maya A and Liam N) and one used Facebook to follow politics (Michael A), the others utilized the medium primarily for socializing with friends and sharing pictures and other social information. In another question Liam N even cites Facebook as an excellent, free resource for promoting one's business, and I think this comment speaks to the fact that while much of Facebook involves socialization, there is great potential for it to be used as an avenue for one's business and profession. This, in turn, puts not only users' personal information but also their personal livelihood at risk if their privacy is violated. Once again, I intend to take the answers to this question – which were fairly straightforward – into consideration when analyzing future answers.

Critiques: Like with Question 1, upon review I believe this may have been better put as a sort of demographic question, as the knowledge gained from it is largely foundational. There was something of a lack of diversity in the answers and beyond what is included above, they tend not to tell much more than face-value. By including them in demographics, I could have potentially shortened the interview and been able to provide more pointed pictures of the participants as a baseline rather than asking what seems to be background information as a full question.

Question 3: What features of Facebook do you make use of? Is Timeline among them?

Despite the nature of the project, I did intend to get at least one user who didn't use timeline for the purpose of comparison. This was also a good chance to ask what sort of information people

were putting up on Facebook in a different manner in the hopes of eliciting responses I might have missed. In the first respect, I did find that Juan C does not have a timeline and that Leonard B may not have one, with his statement having been "Timeline is stupid". Loenard's comment was actually fairly typical, as several of the respondents opted to discuss their opinions on Timeline or how they got it. Of the respondents, Amy C and Lynn V noted being "forced" on to Timeline, and in Lynn V's case she noted it was due to claims that Facebook was forcing users to switch over some few months ago. Having personally switched over to Timeline for the same reason, I can verify this claim, and since then I was informed after the study by Juan C that Facebook finally *did* force users to switch over months after the initial announcement. This notion of people being "forced" on to Timeline, either technically or through misleading announcements, illustrates very clearly one of the key problems of online privacy: users do not always have control of the information they post on websites. This information, according to the respondents, can range from users' photos and complaints about work to users' schedules, calendars, and even their current location down to the address. The nature of the information that can be posted whether the user likes it or not is certainly something researchers should consider when conducting research on "public" information.

Critiques: I have fewer retrospective issues with this question than the last few, given that I seem to have finally escaped the mire of ultimately "demographic" questions. While the Timeline segment of this question could have been a simple demographic query, I do believe the rest of the question – especially in the context of what apps people use rather than what "information" they post – is an important one. Users who do not have full proficiency may not understand the full technicalities of the applications they're using on Facebook, and while I don't think that was an issue with any of my interviewees given the generally limited number of applications and functions they used, when combined with the notion of websites "forcing" changes on people, there should be serious consideration as to how "public" public information really is.

<u>Question 4: Who do you generally intend to see the information you post to Facebook? Do you feel as if you are effective in limiting who can and can't access such information?</u>

This question is one which I feel gets to the heart of the matter: do Facebook users feel like they have control over their own privacy on the website? An overwhelming majority of the respondents intend Facebook to be used for communicating with friends, with Liam N and Maya A also intending it to be frequented by clients. The truly striking aspect of the responses, though, was that out of 10 respondents only 1 (Liam N) did not express some concern over how much control they have over their information. This sentiment expressed itself in multiple forms, ranging from Johnny G's statement that he "hopes" he is effective to Leonard B's statement that he is able to control information "after the fact" or Lynn V's statement that she "tries" to "keep up with" Facebook's security changes. Facebook's management of user privacy seems to be the root of these concerns, with most (6 of 9) of those respondents who were less than sure of their control over security citing Facebook's previous glitches and releases of data or other policies as the reason and the 7th (Lynn V) strongly implying it with the notion that she has to "keep up" with Facebook's security changes. Even Liam N did not praise the privacy settings, but simply didn't speak about them at all and instead focused on Facebook's value as a tool for users. I believe that this notion of a lack of control is very important and may merit a great deal of its

own attention in a continued or separate study. On the surface, I believe the implications of the answers to Question 4 are similar to those of question 3: users do not always intend for their public information to be public, and one interviewee in particular (Michael A) related a story of how his explicitly private and personal information (his e-mail) was made public without his knowledge or release to Facebook.

Critiques: While I do have some critiques of how the question is structured somewhat along the lines of my concerns with the whole survey, I believe the content of this question was vindicated by its effectiveness in eliciting some very striking data from those interviewed. Overall, I think this is one of the questions that I'm most satisfied with and will likely be among the most productive in compiling a case study.

Question 5: How familiar do you feel with Facebook's privacy settings? Do you make use of your own, personal means to control access to Facebook information (super-logout)?

The first thing I garnered from this question is that nobody had even heard of personal means of enhanced privacy protection such as the super-logout or whitewashing one's wall. When pressed about Facebook privacy settings in particular, Leonard B, Liam N, and Sally E all stated that they believed Facebook's privacy settings are sufficient for their needs. Of the rest, though, a once-again common theme in the answer is that while many respondents feel they are familiar with the settings on the surface, Facebook's habit of constantly changing the privacy settings undermines their control over their information. Michael A's specific story of the publication of his e-mail address remains the most striking case of this issue, but the issue is a widespread concern of interviewees across the collected demographics and proficiency levels.

Critiques: While this question might be a bit redundant in the light of the answers to Question 4, I think that being more specific on the matter of Facebook's privacy (given that the study is on Facebook, specifically) has helped shed some light on more specific issues related to the control of one's own personal information.

Question 6: Could the information you or a friend or associate posted to Facebook be a source of harm to you, either personally or professionally?

After much thinking on the answers given to question 6 I've decided to analyze the answers of several respondents individually. While this may be time-consuming, I think the unique stories told by the respondents provide excellent case studies for researchers.

Amy C: Amy C discussed an incident where a friend of hers had her account phished and turned into an outlet for more phishing in turn. Having had it happen to me before and having seen it happen to others, I can verify that these phishing attacks can be very effective and very damaging to a user's account, giving a foreign entity access to all of their personal information regardless of privacy filters by acquiring the account's password. While Facebook has functions which are allegedly intended to prevent this issue, I still personally see likely phishing links from time-to-time, indicating the problem is far from resolved.

Alex Carson
Department of Anthropology
University of Maryland College Park

Liam N: Liam discussed advice he gave to some of his clients to never post anything political on Facebook. They pointed out that he posts political things on his and asked why they can't, to which he replied that being established gave him a sort of leave to do so. While he said it was all about "liability", I think a better term might be "expectations" or "reputation". What I think this story illustrates, though, is that the same sort of info posted by different people can have different impact on the internet just as in the physical world, and I think this is something researchers should take into consideration when conducting their research.

Maya A: While Maya offered no illustrative story, I believe she brought up an important point. In her answer, Maya noted that the release of one's information is "opt-in" instead of "opt-out" in a sense, where users must actively seek to hide their personal information rather than the information being hidden by default (I am aware that the terms are technically backwards, but I couldn't think of a better way of describing them). This goes back to the issues discussed in the case of Google Buzz, where models which release one's information by default often lend themselves to unintentional releases of information and call into question the desires of the users where they are implemented.

Michael A: Michael had a unique insight on this issue. The first possible danger to himself that Michael noted was the fact that he is not out to his family as an LGBT individual. This sort of very personal information could be very harmful to someone if inadvertently released in a way that could be connected to them, and I believe Michael's example is a very good reminder of some of the stakes involved in protecting privacy.

Of the other respondents, both Susan D and Sally E remarked on photo-tagging – a practice in which an individual "tags" someone as being in a photo, potentially making it viewable to people off the tagged user's friend list and outside of their privacy allowances – being a problem both of them have had experience with. While users are theoretically informed of photo tagging and tags can be removed, the problem of the permanence of digital information as discussed ad nauseum becomes a significant problem for users here. Of all of the respondents, Juan C was not concerned over the possibility due to a perceived lack of harmful information on his wall while Leonard B was confident in his security.

Critiques: The most obvious critique on this question and my analysis is that I come from a significantly biased perspective on this issue, but it seemed that a number of my respondents shared the same concerns in one way or another. While I seem to have focused overwhelmingly on the problematic stories told, I did so because I think these provide apt companions for the case studies in the first half of this report and will provide valuable data for the IRB.

<u>Question 7: Would there be negative consequences for you or someone you know if information on your Facebook was accessed by someone who was not intended to be able to?</u>

In general, the respondents agreed that there would be negative consequences, even if it was only to their personal dignity and peace of mind. Several agreed that someone taking their personal information was the most apparent risk in their eyes, particularly Liam N in reference to his clients. Of all the respondents, Sally E had the most detailed answers to this question, discussing

it over the course of two personal anecdotes. The first was an enlightening story about how someone she knew found a website about her written by someone with less than positive opinions. Another story was about how another friend had posted a rather innocuous comment on Facebook about not feeling like going to work, only for her manager to find it and fire her due to the comment. I think Sally's first story in particular may be very important, as it brings to the forefront the notion that not all information publicly available on a person is written by them or with their leave. Human beings can say things about each other on the internet even more effectively and permanently than in strictly physical space, and this raises some challenges for researchers when it comes to the ethics of using such information.

Critiques: Once again, I do admit to a bias on this issue in favor of privacy, but the answers here were rather commonplace and straightforward. I think this may be in part because most of the issues addressed here were discussed by respondents answering to the previous question, but I think Sally E's contributions definitely validate the question as a whole.

Question 8: Do you think that researchers should be able to use publicly-available information from a Facebook account without acquiring the consent of the account's owner?

This question was relatively straightforward, but the responses show quite a bit of variety. Lynn V, Juan C, Johnny G, and Leonard B gave rather open-and-closed "No" answers to this question, where Amy C and Sally E both stated that researchers should err on the side of ethics and generally assume that they should ask consent for any information they take from someone. Maya A had different standards for information, stating that information such as "whether someone has a profile picture" could simply be taken but stating that any personal information taken from a website should require consent. Michael A believed that due to Facebook's current policies related to privacy, researchers should ask before taking any information because Facebook's changing standards can release information one would prefer not be. Susan D and Lian N both expressed that they believed this practice would happen whether they liked it or not, with Liam N noting the financial value of information and Susan D qualifying that she would prefer that people be asked by researchers before public information is taken.

Critiques: I have intentionally foregone in-depth analysis of the answers to this question because I believe the researchers who read the case study that this project produces should decide for themselves what it all means. My personal perspective from my own experience – especially after doing this project – is that with the current state of the internet researchers should assume absolutely nothing has been made private or public by intent, and that research participants should be involved to the greatest degree possible in what information of theirs is used. In compiling a case study for this data, though, I encourage researchers to come to their own conclusion based on the anthropological research and current events when it comes to protecting their participants.

Question 9: Do you have anything else you'd like to add?

Many respondents declined to answer this question outright, where most of the others felt that they had nothing to add. Of those who did, however, they tended to share their own insights on

Alex Carson
Department of Anthropology
University of Maryland College Park

Facebook. Amy C commented on the likelihood that social networking was around to stay, and noted that as long as people remained ignorant of how social networking websites function they will be taken advantage of. Leonard B remarked that he was excited to see what would come after Facebook, stating that if Facebook failed to protect its users' privacy someone else would come along. Liam commented on Facebook's value as a business tool and the resources out there to help businesses rendering a traditional website essentially obsolete. All in all, I believe that these are rather self-explanatory insights that, nonetheless, give us things to think about as researchers continue to form an ethical framework for their research on the internet.