



OPINION ARTICLE

Ethical research standards in a world of big data [v1; ref status: approved with reservations 1, <http://f1000r.es/2wqj>]

Caitlin M. Rivers, Bryan L. Lewis

Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech., Blacksburg, VA, 24060, USA


v1 First published: 06 Feb 2014, 3:38 (doi: [10.12688/f1000research.3-38.v1](https://doi.org/10.12688/f1000research.3-38.v1))
Latest published: 06 Feb 2014, 3:38 (doi: [10.12688/f1000research.3-38.v1](https://doi.org/10.12688/f1000research.3-38.v1))

Abstract

In 2009 Ginsberg *et al.* reported using Google search query volume to estimate influenza activity in advance of traditional methodologies. It was a groundbreaking example of digital disease detection, and it still remains illustrative of the power of gathering data from the internet for important research. In recent years, the methodologies have been extended to include new topics and data sources; Twitter in particular has been used for surveillance of influenza-like-illnesses, political sentiments, and even behavioral risk factors like sentiments about childhood vaccination programs. As the research landscape continuously changes, the protection of human subjects in online research needs to keep pace. Here we propose a number of guidelines for ensuring that the work done by digital researchers is supported by ethical-use principles. Our proposed guidelines include: 1) Study designs using Twitter-derived data should be transparent and readily available to the public. 2) The context in which a tweet is sent should be respected by researchers. 3) All data that could be used to identify tweet authors, including geolocations, should be secured. 4) No information collected from Twitter should be used to procure more data about tweet authors from other sources. 5) Study designs that require data collection from a few individuals rather than aggregate analysis require Institutional Review Board (IRB) approval. 6) Researchers should adhere to a user's attempt to control his or her data by respecting privacy settings. As researchers, we believe that a discourse within the research community is needed to ensure protection of research subjects. These guidelines are offered to help start this discourse and to lay the foundations for the ethical use of Twitter data.

Article Status Summary

Referee Responses

Referees	1
v1 published 06 Feb 2014	 report

1 Tristan Henderson, University of St Andrews UK

Latest Comments

Ernesto Priego, City University London, UK
27 May 2014 (V1)

Corresponding author: Caitlin M. Rivers (cmrivers@vbi.vt.edu)

How to cite this article: Rivers CM and Lewis BL (2014) Ethical research standards in a world of big data [v1; ref status: approved with reservations 1, <http://f1000r.es/2wqj>] F1000Research 2014, 3:38 (doi: [10.12688/f1000research.3-38.v1](https://doi.org/10.12688/f1000research.3-38.v1))

Copyright: © 2014 Rivers CM and Lewis BL. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: Research reported in this publication was supported by the US National Institute of General Medical Sciences of the National Institutes of Health under award number 2U01GM070694-09. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support provided by Defense Threat Reduction Agency Validation Grant HDTRA1-11-1-0016.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 06 Feb 2014, 3:38 (doi: [10.12688/f1000research.3-38.v1](https://doi.org/10.12688/f1000research.3-38.v1))

Introduction

The growing popularity of social media sites presents a unique opportunity to study human interactions and experiences. Twitter, one of the most popular social media sites, allows users to ‘micro-blog’ by sharing 140 character messages with their social network. Although Twitter doesn’t disclose the number of people who use its service, estimates are in the hundreds of millions and perhaps as many as half a billion¹. Approximately 340 million tweets are sent every day around the world². Researchers have begun to use this data to answer questions in a variety of fields³⁻⁹. Recent reflections on the data collection practices of the US National Security Administration have spurred similar meditations on the ethics of digital research¹⁰. The concern is that Twitter data could conceivably be used in a way that violates the privacy and rights of the tweet authors.

Twitter data have already been used in a number of studies to detect influenza-like illness (ILI)³⁻⁶, risky behaviors associated with the transmission of HIV⁷, sentiments about childhood vaccination programs⁸, and political sentiments⁹. These study designs generally feature count data, rather than user-specific data. For example, there are multiple studies that compare the proportion of tweets about flu-related symptoms to public health data on influenza-like incidence (ILI). An increase in syndromic flu tweets might indicate that an outbreak is occurring. These data are usually reported either at the national level or without any geographic parameters. Another common study design aims to determine public sentiments by counting words, phrases, and emoticons that co-occur with keywords like ‘Obama’. These sentiment indicators can be used to infer public opinions about political elections, mental health, or consumer products.

Unlike Facebook, Pinterest and other competitors, Twitter provides several application programming interfaces (APIs) that allow

real-time access to vast amounts of content. Data streamed through the APIs include metadata about the authors, including the text location from their profile (e.g. ‘Baltimore’), their time zone, the time they sent the tweet, the number of friends and followers they have, the number of tweets they have ever sent, and more (Figure 1). Approximately 1% of tweets have a geolocation, which uses GPS to append the author’s precise geographic coordinates to the tweet. Geolocations are sufficiently detailed to determine from which wing of a building a tweet was sent. The default privacy settings do not enable geolocation, but do make a user’s tweets and metadata available through the API. Users can modify their setting to make their profile private, which shields their account from public view online and from the API¹¹.

There are numerous ways to access the data through these APIs: one method is through the ‘garden hose’, which is a random sample of approximately 1% of all live-streamed tweets. Other access methods include the search API which enables searching for particular users, hashtags, or locations, and author-specific queries which can retrospectively gather up to 3,200 tweets from a single user¹¹. Furthermore, in 2010 Twitter donated its entire historical record of tweets to the US Library of Congress. Detailed plans for these data are not yet available, but the Library of Congress has indicated that it intends to collaborate with academic institutions to make the data available to researchers¹².

The strength of tweets as a data source is in the volume; collection through the garden hose API brings in approximately 60,000–100,000 tweets per day. However because tweets are short and often lack context, it is difficult for computers to determine tweet content automatically. For this reason, researchers primarily use tweet data to conduct population-level research concerned with trends and

```

27.     "in_reply_to_user_id_str": null,
28.     "contributors": null,
29.     "text": "Introducing the Twitter Certified Products Program: https://t.co/MjJ8xAnT",
30.     "retweet_count": 121,
31.     "in_reply_to_status_id_str": null,
32.     "id": 240859602684612608,
33.     "geo": null,
34.     "retweeted": false,
35.     "possibly_sensitive": false,
36.     "in_reply_to_user_id": null,
37.     "place": null,
38.     "user": {
39.         "profile_sidebar_fill_color": "DDEEF6",
40.         "profile_sidebar_border_color": "C0DEED",
41.         "profile_background_tile": false,
42.         "name": "Twitter API",
43.         "profile_image_url":
44.         "http://a0.twimg.com/profile_images/2284174872/7df3h38zabcvjylnyfe3_normal.png",
45.         "created_at": "Wed May 23 06:01:13 +0000 2007",
46.         "location": "San Francisco, CA",
47.         "follow_request_sent": false,
48.         "profile_link_color": "0084B4",
49.         "is_translator": false,
49.         "id_str": "6253282",

```

Figure 1. An excerpt of a single tweet returned through the Twitter API¹¹.

patterns. Study designs rely on large volumes of data to accommodate false positives and negatives. A typical data set contains millions of tweets and many thousands of tweet authors. However, a user-centric use case involving Twitter is not inconceivable. Researchers interested in social network analysis, qualitative research, and rare-event topics may eventually turn to Twitter as a data source. Potential methodologies include building a social network out of @mentions (the @ is Twitter lexicon for referencing another user); mining qualitative data from specific user's accounts; or conducting prospective research by following a person or small group of people over time. These user-centric approaches are fundamentally different from population-level studies, and may require different ethical considerations than aggregated study designs. Additional methodologies might also involve interacting with Twitter users, which will not be addressed here.

Under non-digital circumstances, ethics guidelines suggest that collecting information from a public space where people could 'reasonably expect to be observed by strangers' is considered appropriate

even without informed consent¹³. According to these guidelines, tweets are text that users publish for the purpose of sharing with others. The weakness of this argument is that it fails to distinguish between population-level research and research focused on selected individuals. It would be clearly unethical for a researcher to follow one specific shopper around the mall and gather data exclusively about him without his consent. However, simply counting or observing behavior in aggregate at a mall is an acceptable research practice. The difference is that the latter example adheres to a level of privacy that the observed individual might expect from being in public, whereas the former violates those natural privacy boundaries. A similar distinction is needed in digital research.

As an example of the potential privacy pitfalls of digital research, suppose investigators were interested in the social networks of adolescents suffering from depression. A research plan might look like this: the investigators gather geocoded tweets that contain words relevant to the topic of interest, as shown in Figure 2. They filter for geocodes that correspond to school locations in order to

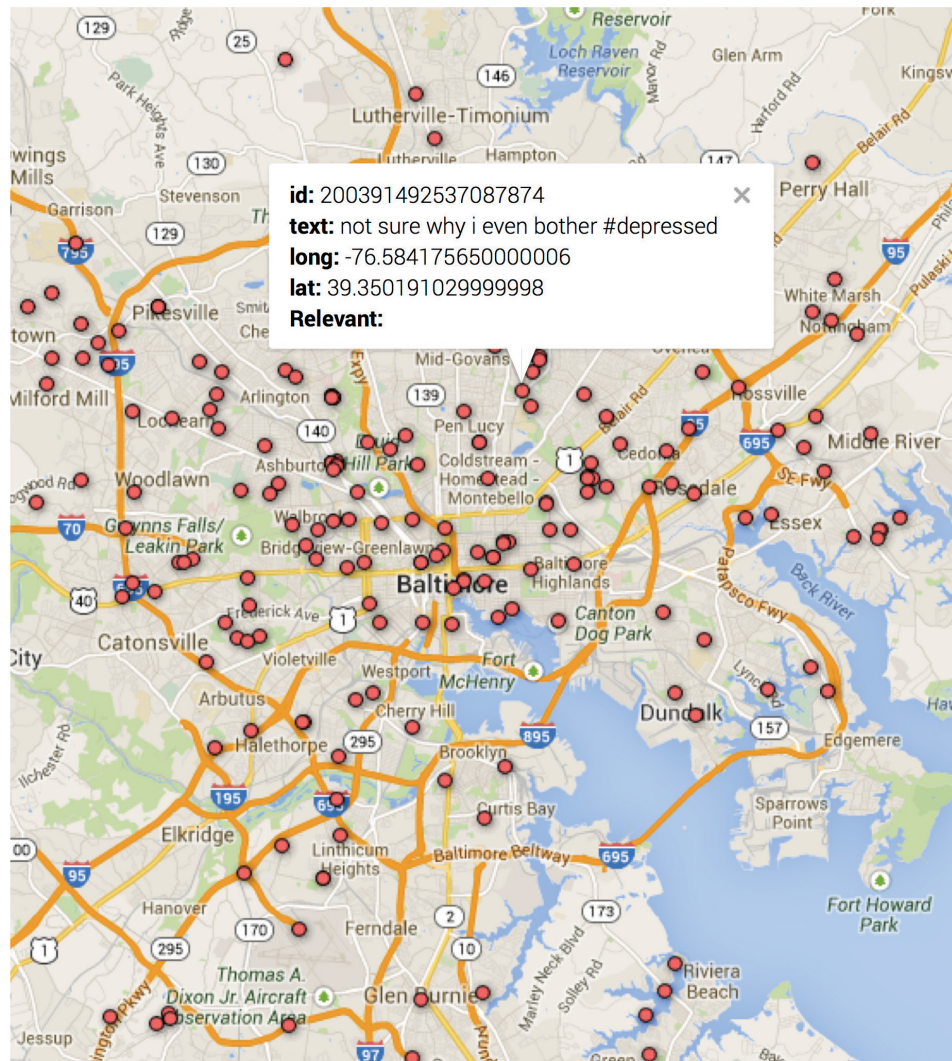


Figure 2. Each dot is a geolocated tweet collected through the Twitter API. The example tweet displayed is fabricated.

identify adolescent users. From there, a simple query to the Twitter API returns a list of followers for each of those presumably depressed adolescents. They now have a social network. For each member of the network, they mine the user's tweet histories to find identifying details such as their real names. The researchers then use the gathered information to 'snowball' data collection by curating from a variety of different sources like Facebook, tumblr and the White Pages. They can collect birth dates, cell phone numbers, home addresses, favorite hangout spots, "likes" and "dislikes", etc. The final result would be detailed demographic information for potentially thousands of people who exhibit symptoms of depression or are connected to a depressed adolescent. Current guidelines do not prohibit this kind of research activity. However, if the same information were collected through surveys or other traditional means, Institutional Review Board (IRB) approval would be needed.

According to the US Department of Health and Human Services Policy for Protection of Human Research Subjects, data that are publicly available are exempt from requiring IRB approval¹⁴. Because Twitter data are public, they technically fall under this exemption. Furthermore, Twitter's privacy policy makes no secret of the fact that user data are indexed by search engines, archived within the US Library of Congress, and are available through an API¹⁵. However, it is unlikely that many users follow the link to read the lengthy and complex document. One study found that it would take 244 hours a year for an average internet user to read every privacy policy of the unique sites they visit¹⁶.

The US Consumer Privacy Bill of Rights (CPBR)¹⁷ may instead serve as a useful framework for guiding researchers conducting research with Twitter. The CPBR was issued by the Obama administration in February 2012 in order to "give consumers clear guidance on what they should expect from those who handle their personal information, and set expectations for companies that use personal data." There are seven principles enumerated by CPBR in [Table 1](#).

Proposed guidelines for the ethical use of Twitter data **The objectives, methodologies, and data handling practices of the project are transparent and easily accessible**

This information should be published in manuscripts, published on the web for the public to access, and provided to IRB (when relevant). Going forward, collaboration between the research community and Twitter to provide information to users about ongoing

research and relevant results may also be beneficial. Transparency regarding uses of Internet data for research purposes is needed for fostering 'privacy literacy' so that the users can make informed decisions about participating in Twitter.

Study design and analyses respect the context in which a tweet was sent

A tweet author discussing his mental health, for example, does not do so with the intention of sharing that data with researchers; he does it to communicate with his digital community. Qualitatively analyzing these communications as if they are offered for research consumption does not align with the context in which the tweets were created. Twitter participants can reasonably expect to rely on some anonymity of the crowd to manage privacy.

The anonymity of tweet authors is protected, ensuring that subjects should not be identifiable in any way

To preserve source anonymity, direct quotes or screen names are not publishable, nor are any details that could be used to identify a subject. Any and all information that could be entered into a search engine to trace back to a human source should be protected. A composite of multiple example tweets may instead be used for illustrative purpose. Geolocations in particular should be scaled to a larger geographic area in order to avoid violating the privacy of those tweet authors. The Title 13 of the Data Protection and Privacy Policy, the federal law under which the Census Bureau is regulated, expressly forbids publishing GPS coordinates¹⁸; researchers should adhere to this guideline as well.

Tweet data are not used to harvest additional information from other sources

Focused collection is also important for preserving anonymity. It is possible to use data collected from Twitter to discern the identities of tweet authors, which can then be used to find and collect additional information from additional sources. For example an author's username, identifying details provided in tweet texts, or geolocations could all be used to collect data about that individual from other sources like Facebook, LinkedIn, Flickr, or public records.

Twitter users' efforts to control their personal data are honored

Researchers may not follow a user on Twitter in order to gain access to a protected account. Doing so would violate that user's efforts to control his or her personal data.

Table 1. Consumer Privacy Bill of Rights¹⁷.

Transparency	Consumers have a right to easily understandable and accessible information about privacy and security practices.
Respect for context	Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.
Security	Consumers have a right to secure and responsible handling of personal data.
Focused collection	Consumers have a right to reasonable limits on the personal data that companies collect and retain.
Accountability	Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.
Individual control	Consumers have a right to exercise control over what personal data companies collect from them.
Access and accuracy	Consumers have a right to access and correct personal data in usable formats.

Researchers work collaboratively with IRB just as they would for any other human subject data collection

There is not currently an expectation that researchers engaging in research using Twitter will interface with their IRB. As discussed above, studies that could be conceived as individual-based should require IRB approval, whereas research designs that use data in aggregate (e.g. counts of keywords) may proceed without explicit consent. In turn, review boards should keep abreast of social network mining methodologies and corresponding ethical considerations in order provide informed guidance to researchers.

Conclusions

Research involving Twitter is growing in popularity, but the issues surrounding the ethics of using it as a data source have not yet been closely examined. There are hypothetical study designs that could use Twitter data in a way that violates the privacy and ethical treatment of participants. In order to avoid those misuses, six guidelines derived from the US Consumer Privacy Bill of Rights are proposed. We welcome discourse in the research community on this topic, and encourage further discussion.

Please use the #EthicalTwitter hashtag on Twitter to participate in this discussion online.

Author contributions

CR conceived of the guidelines and drafted the manuscript. BL contributed to both the development of the guidelines and preparation of the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

Research reported in this publication was supported by the US National Institute of General Medical Sciences of the National Institutes of Health under award number 2U01GM070694-09. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support provided by Defense Threat Reduction Agency Validation Grant HDTRA1-11-1-0016.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments.

References

1. **Twitter Reaches Half a Billion Accounts, More than 140 Millions in the U.S.** *SemioCast*. 2012.
[Reference Source](#)
2. **Twitter turns six.** *Twitter Blog*. 2012.
[Reference Source](#)
3. Ginsberg J, Mohebbi M, Patel R, *et al.*: **Detecting Influenza Epidemics Using Search Engine Query Data.** *Nature*. 2009; **457**(7232): 1012–1014.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Achrekar H, Gandhe A, Lazarus R, *et al.*: **Predicting Flu Trends using Twitter Data.** *The First International Workshop on Cyber-Physical Networking Systems*. 2011; 713–718.
[Reference Source](#)
5. Lamos V, Cristianini N: **Tracking the flu pandemic by monitoring the social web.** *2nd IAPR Workshop on Cognitive Information Processing*. 2010; 411–416.
[Publisher Full Text](#)
6. Salathé M, Bengtsson L, Bodnar TJ, *et al.*: **Digital Epidemiology.** *PLoS Comput Biol*. 2012; **8**(7): e1002616.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Young S, Rivers C, Lewis B: **Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes.** *Preventative Medicine*. 2014.
8. Campbell E, Salathé M: **Complex Social Contagion Makes Networks More Vulnerable to Disease Outbreaks.** *Sci Rep*. 2013; **3**: 1905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Tumasjan A, Sprenger TO, Sandner PG, *et al.*: **Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.** *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 2012; 178–185.
[Reference Source](#)
10. Omand D: **NSA leaks: how to make surveillance both ethical and effective.** *The Guardian*. 2013.
[Reference Source](#)
11. **Twitter Documentation.** *Twitter*. 2013.
[Reference Source](#)
12. Raymond M: **How Tweet It Is!: Library Acquires Entire Twitter Archive.** *Library of Congress Blog*. 2010.
[Reference Source](#)
13. The British Psychological Society. **Report of the Working Party on Conducting Research on the Internet.** *Ethics*. 2007; **23**(4): 257–257.
[Reference Source](#)
14. **Code of federal regulation: Protection of human subjects.** *Department of Health and Human Services*. 2009.
[Reference Source](#)
15. **Twitter Privacy Policy.** *Twitter*. 2013.
[Reference Source](#)
16. McDonald AM, Cranor LF: **The Cost of Reading Privacy Policies.** *I/S: A Journal of Law and Policy for the Information Society*. 2008.
[Reference Source](#)
17. **Consumer Privacy Bill of Rights.** 2012.
[Reference Source](#)
18. **Data Protection and Privacy Policy.** *U.S. Census Bureau*. 2012.
[Reference Source](#)

Current Referee Status:

Referee Responses for Version 1



Tristan Henderson

Department of Computer Science, University of St Andrews, Fife, UK

Approved with reservations: 06 March 2014

Referee Report: 06 March 2014

This opinion article is interesting and timely, as it is clear that the quantity of research using social media and other sources of personal data is increasing, and the potential concerns around this need to be considered by researchers and ethics committees. I found the current form of the paper to be rather unsatisfactory, however: it is not clear how or where the proposed guidelines would be applied, how they were developed, and most importantly there is no contextualisation of how this paper builds on the myriad discussions that have already taken place in this arena.

The introductory section of the paper is fine, and provides a clear description of how Twitter works, which is no doubt of use for the intended audience. But the authors need to clearly describe the purpose of the guidelines. Indeed there needs to be some introductory text before the guidelines are introduced, as the current text is rather incongruent. Is the text on the Consumer Privacy Bill of Rights supposed to be in a separate section following the introduction?

Moreover the aim and applicability of the guidelines is unclear. How would these guidelines be enforced? Would they be applied to all studies? Is there any mechanism for exceptions to this? For instance if I wanted to do a study comparing health behaviour on Twitter and Facebook, I would need to collect data from both networks. But this violates the guidelines. Or are these guidelines only for studies where IRB approval would not be obtained? There is much discussion over whether social media studies require any IRB approval at all ^{1,2,3}. But your last guideline seems to disagree with this. The title of the paper refers to “big data” but the paper itself really concerns itself with Twitter (with some discussion of linkability). Are the guidelines to apply to all big data studies?

Given that this paper really only targets US ethics committees and IRBs, I was surprised that there was no discussion of the Common Rule or the recent Menlo Report ^{4,5}. An opinion article in Science last year also discussed US regulation ⁶.

One of the guidelines explicitly describes the context in which a tweet was sent. But there is no mention of contextual integrity ⁷ which seems to be the most common framework for studying this, and has indeed been applied to social media research ^{8,9,10}.

Danah Boyd has talked widely about the ethics of big data research ^{11,12} and her work should be discussed. Concerns about Twitter specifically have also been raised ^{13,14} and Neuhaus and Webmoor specifically propose “agile ethics” for the study of such networks ¹⁵. How would agile ethics fit in with your guidelines?

The implicit assumption in your guidelines is that studies should occur without consent and involve trawling the firehose. But these kinds of studies in themselves are controversial ¹⁶. Moreover is there any reason why informed consent could not be obtained for some of these studies, especially since you touch on consent when discussing your hypothetical shopping centre study? We have looked at this empirically ¹⁷.

One striking difference between research and the collection of data for business purposes is that researchers are typically interested in reproducibility. Indeed many research funding bodies worldwide are now insisting on data archiving and sharing as a requirement, and this would apply to social media studies as well (something that has been criticised in the past ¹⁸). Should your guidelines not address data sharing? Technical challenges exist, including the linkability problems that you discuss ¹⁹ and the difficulty of anonymising data ²⁰, while solutions are still nascent ^{21,22}.

The BPS reference (13) should be updated ²³. Moreover you ought to compare and contrast other ethical guidelines for Internet research ^{24,25,26}, and other proposals for social media and health research ^{27,28}.

At one point you mention “privacy literacy” but there is no reference and it was unclear what you mean here. Privacy “salience” has been discussed in the literature ²⁹ but might be slightly different.

Finally, I qualify my review by stating that I am not a life scientist and that this is the first review that I have written for F1000Research. Indeed, I had not heard of it prior to my review request. I do question the appropriateness of a scientific journal for this work. If the intention of the authors is, as stated, to encourage more discussion then it would be good to focus this in one place. Since F1000Research allows comments from readers, then perhaps this does provide a better avenue for discussion than over Twitter, which can be difficult over long time-scales and is not particularly persistent. But in my opinion an even better mechanism for discussion and debate would be to hold a workshop and then publish a workshop report. Certainly further discussion with other communities, and examination of the discussion that has already taken place, would be useful to inform and improve this document. In the first instance, I think it would be good to see a revised document that shows more awareness of the current work, and where controversy exists, discusses why the authors have taken their chosen stance when proposing their guidelines.

References

1. Buchanan E, Aycock J, Dexter S, Dittrich D, Hvizdak E: Computer Science Security Research and Human Subjects: Emerging Considerations for Research Ethics Boards. *Journal of Empirical Research on Human Ethics*. 2011; **6** (2): 71-83 [Publisher Full Text](#)
2. Halavais A: Social science: Open up online research. *Nature*. 2011; **480** (7376): 174-175 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Solberg LB: Data Mining on Facebook: A free space for researchers or an IRB nightmare?. *University of Illinois Journal of Law, Technology & Policy*. 2010; **2010** (2): 311-343 [Reference Source](#)
4. Dittrich D, Kenneally E: The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *Cooperative Association for Internet Data Analysis*. 2012. [Reference Source](#)
5. Bailey M, Dittrich D, Kenneally E, Maughan D: The Menlo Report. *IEEE Security & Privacy Magazine*. 2012; **10** (2): 71-75 [Publisher Full Text](#)
6. Shapiro RB, Ossorio PN: Research Ethics. Regulation of online social network studies. *Science*. 2013; **339** (6116): 144-145 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Nissenbaum HF: Privacy as Contextual integrity. *Washington Law Review*. 2004; **79** (1): 119-157

Reference Source

8. Page X: Contextual integrity and preserving relationship boundaries in location- sharing social media In Proceedings of the ACM CSCW Workshop on Measuring Networked Social Privacy: Qualitative & Quantitative Approaches, San Antonio, TX, USA, February. 2013. [Reference Source](#)
9. Shi P, Xu H, Chen Y: Using contextual integrity to examine interpersonal information boundary on social network sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France*. 2013. 35-38 [Publisher Full Text](#)
10. Hull G, Lipford HR, Latulipe C: Contextual gaps: privacy issues on Facebook. *Ethics Inf Technol*. 2011; **13** (4): 289-302 [Publisher Full Text](#)
11. Boyd D: Privacy and publicity in the context of big data. Keynote at WWW '10: the 19th International Conference on the World Wide Web, April. 2010. [Reference Source](#)
12. Boyd D, Crawford K: Critical questions for big data. *Inf Commun Soc*. 2012; **15** (5): 662-679 [Publisher Full Text](#)
13. Puschmann C, Burgess J: The politics of Twitter data. Social Science Research Network Working Paper Series, January. 2013. [Reference Source](#)
14. Vieweg S: The ethics of Twitter research. *Revisiting Research Ethics in the Facebook Era: Challenges in Emerging CSCW Research, Savannah, GA, USA, February*. 2010. [Reference Source](#)
15. Neuhaus F, Webmoor T: Agile ethics for massified research and visualization. *Inf Commun Soc*. 2012; **15** (1): 43-65 [Publisher Full Text](#)
16. Ioannidis JPA: Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth*. 2013; **13** (4): 40-42 [PubMed Abstract](#) | [Publisher Full Text](#)
17. McNeilly S, Hutton L, Henderson T: Understanding ethical concerns in social media privacy studies. In Proceedings of the ACM CSCW Workshop on Measuring Networked Social Privacy: Qualitative & Quantitative Approaches, San Antonio, TX, USA, February. 2013. [Reference Source](#)
18. Huberman BA: Sociology of science: Big data deserve a bigger audience. *Nature*. 2012; **482** (7385): 308 [PubMed Abstract](#) | [Publisher Full Text](#)
19. Narayanan A, Shmatikov V: De-anonymizing social networks. *Proceedings of the IEEE Symposium on Security and Privacy*. 2009. 173-187 [Publisher Full Text](#)
20. Zimmer M: "But the data is already public": on the ethics of research in Facebook. *Ethics Inf Technol*. 2010; **12** (4): 313-325 [Publisher Full Text](#)
21. Dwork C: Differential Privacy. *M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, editors, Automata, Languages and Programming, volume 4052 of Lecture Notes in Computer Science, Springer, Berlin*. 2006. 1-12 [Publisher Full Text](#)
22. Hutton L, Henderson T: An architecture for ethical and privacy-sensitive social network experiments. *SIGMETRICS Perform. Eval. Rev*. 2013; **40** (4): 90-95 [Publisher Full Text](#)
23. British Psychological Society: Ethics guidelines for Internet-mediated research. 2013. [Reference Source](#)
24. Buchanan EA, Ess C: Internet research ethics: The field and its critical issues. *The Handbook of Information and Computer Ethics, number 11*. 2008. 273-292 [Publisher Full Text](#)
25. Eynon R, Schroeder R, Fry J: New techniques in online research: challenges for research ethics. *Twenty-first Century Society*. 2009; **4** (2): 187-199 [Publisher Full Text](#)
26. Felzmann H: Ethical issues in internet research: international good practice and Irish research ethics documents. In C. Fowley, C. English, and S. Thouësnny, editors, Internet research, theory, and practice: perspectives from Ireland, pages 11–32. Research-publishing.net, Dublin, 1 edition. 2013. [Reference Source](#)
27. McKee R: Ethical issues in using social media for health and health care research. *Health Policy*. 2013; **110** (2-3): 298-301 [PubMed Abstract](#) | [Publisher Full Text](#)
28. Moreno MA, Fost NC, Christakis DA: Research ethics in the MySpace era. *pediatrics*. 2008; **121** (1): 157-161doi [PubMed Abstract](#)

29. Bonneau J, Preibusch S: The privacy jungle: On the market for data protection in social networks. *Proceedings of The Eighth Workshop on the Economics of Information Security (WEIS 2009)*. 2009. 121-167

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Article Comments

Comments for Version 1

Ernesto Priego, City University London, UK

Posted: 27 May 2014

Inspired by this piece and the SA post that indirectly led me to it, I wrote this post: [Twitter as Public Evidence and the Ethics of Twitter Research](#) (27/05/14).

Competing Interests: None.
