Ethics and online social network research – developing best practices

Tristan Henderson
School of Computer Science
University of St Andrews
St Andrews, UK
http://www.cs.st-andrews.ac.uk/~tristan/
tnhh@st-andrews.ac.uk

Luke Hutton
School of Computer Science
University of St Andrews
St Andrews, UK
http://www.cs.st-andrews.ac.uk/~lhutton/
lh49@st-andrews.ac.uk

Sam McNeilly
School of Computer Science
University of St Andrews
St Andrews, UK
sm2269@st-andrews.ac.uk

Social network sites (SNSs) and other online social networks such as Facebook and Twitter represent a huge source of data for research in many fields, including sociology, medicine, anthropology, politics and computer science. Such sites may contain sensitive information and care needs to be taken when designing experiments or collecting SNS data. We outline some of the potential ethical concerns, describe our efforts to develop best practices, and solicit help with outstanding challenges.

ethics, social network sites, research

1. INTRODUCTION

Facebook and other social network sites (SNSs) are used by hundreds of millions of people daily.[1] With such a large number of social interactions being made and recorded digitally, it is not surprising that researchers from many fields in the humanities and both physical and social sciences have exploited this rich source of data, with one recent survey listing 410 social science papers studying the Facebook SNS alone (Wilson et al. 2012). This includes HCI researchers, who have studied, for instance, mobile SNSs to understand location-sharing (Barkhuus et al. 2008) behaviour, or privacy preferences (Toch et al. 2010). But there are many ethical issues that need to be considered when dealing with SNS data. Is it the case that data published to an SNS are truly "public" and fair game for researchers? Are there legal requirements such as Data Protection concerns? Should SNS participants provide informed consent?

Some SNS studies have raised methodological concerns about ethics and privacy (Zimmer 2010). Our interests are two-fold. First, we wish to understand the potential ethical concerns of SNS research. Second, we would like to collect, develop and distribute best practices for conducting such research, thus allowing the research community to leverage the vast amount of SNS data while minimising harm to SNS users.

We are developing an architecture that is designed to enable privacy-sensitive SNS experiments. But several challenges remain to making this architecture a useful tool for researchers, and we hope to discuss these at this workshop.

2. ETHICS AND SNS RESEARCH

Ethical concerns with SNS research are various and have begun to be discussed by the research community. Neuhaus and Webmoor propose "agile ethics" for academic researchers, who they argue need to take more care with SNS data than commercial SNS providers themselves (Neuhaus & Webmoor 2012). Moreno et al. examine the conduct of research with adolescents using SNSs, and determine that informed consent may be required but should be decided on a case-by-case basis, and that the consent of the SNS community and operator are not required (Moreno et al. 2008). Zimmer outlines the problems with a well-known Facebook study conducted at Harvard (Lewis et al. 2008) and concludes that just because a user shares data on an SNS, this does not mean that a researcher can collect these shared data; on the contrary, researchers must take even more care with such data (Zimmer 2010).

[1] http://newsroom.fb.com/content/default.aspx?NewsAreaId=22

A social network site by its very nature comprises connections between many users, and many concerns arise because of the number of, and connections between, the various key actors. These include the SNS users participating in an experiment, their friends (who may be mentioned or included in participants' data), other SNS users with whom the participants may have shared data and the operators of the SNS. The researchers themselves may also be considered key actors, as might any other researchers with whom data might be shared. Each of these actors will have different concerns about the SNS data that are generated and collected. A participant might be willing to share some data with a researcher, but not other data. A participant's friend might be completely unwilling to have their data shared. Conversely, a researcher's goal might well be to collect as much data as possible. How can we address the tension between these concerns? Can we encode this tension into our experimental design?

3. OUR APPROACH

We are developing an architecture for enabling privacy-sensitive and ethical SNS experiments. Further details of our system, entitled PRISONER (Privacy-Respecting Infrastructure for Social Online Network Experimental Research) can be found in (Hutton & Henderson 2012). Central to our architecture is the notion of an *experimental policy*, which captures the privacy policy and ethics requirements of a given experiment. A researcher wishing to run an SNS experiment will first design such a policy in XML (e.g., Figure 1). The policy expresses what SNS information can be collected from experimental participants and their SNS friends, how the data will be stored and/or sanitised. The policy is then submitted to our system, and all attempts to collect and use SNS data also go through the system. PRISONER enforces the policy and prevents accidental capture or use of SNS data (Figures 2 and 3). We have developed a common abstraction, termed *social activity clients* to allow SNS data to be collected from a variety of SNSs, e.g., Facebook, Twitter and last.fm, and also other non-SNS sources of networked interactions such as citation networks. We have also developed infrastructure for what we term *participation clients* for conducting experiments, such as web-based questionnaires, and experience sampling method software for smartphones.

In some senses our aims are similar to attempts to embed Cavoukian's "privacy by design" principles (Cavoukian 2011) into managing experiments (Office of the Information & Privacy Commissioner of Ontario & Children's Hospital of Eastern

```
<?xml version="1.0" encoding="UTF-8"?>

<p:privacy-policy
  xmlns:p="http://prisoner.cs.st-andrews.ac.uk/prisoner/privacy-policy"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://prisoner.cs.st-andrews.ac.uk/prisoner/privacy-policy
    privacy_policy.xsd">

  <policy for="Facebook:User">
    <attributes>
      <attribute type="displayName">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="image">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="birthday">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="gender">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="interestedIn">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="relationshipStatus">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="significantOther">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="politicalViews">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
      <attribute type="religion">
        <attribute-policy allow="retrieve"></attribute-policy>
      </attribute>
    </attributes>

    <object-policy allow="retrieve">
      <object-criteria>
        <and-match>
          <attribute-match match="author.id" on_object="session:Facebook.id" />
        </and-match>
      </object-criteria>
    </object-policy>
  </policy>

  <policy for="Facebook:Friends">
    <object-policy allow="retrieve">
      <object-criteria>
        <and-match>
          <attribute-match match="author.id" on_object="session:Facebook.id" />
        </and-match>
      </object-criteria>
    </object-policy>
  </policy>
</p:privacy-policy>
```

Figure 1: *An example experimental policy, allowing an researcher to collect and store selective Facebook data.*

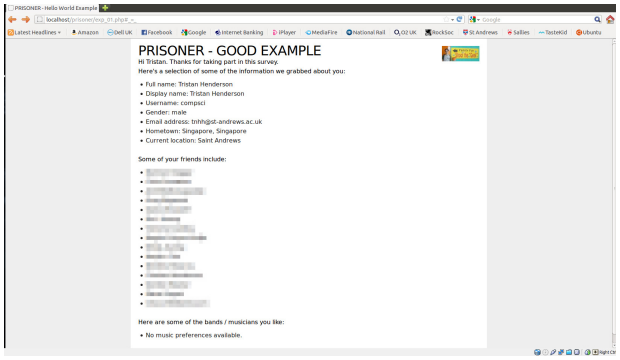

Figure 2: *PRISONER provides a common interface for accessing social network site data, and can be easily incorporated into participation clients such as this web site.*

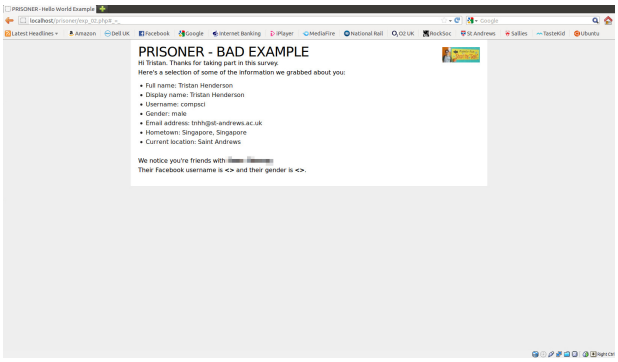

Figure 3: *An inadvertent attempt to capture more data than is allowed by the experimental policy (in this case, data about friends of a participant) is prevented by our PRISONER system.*

Ontario 2011), but we go further in that we wish to embed thinking about privacy into experimental design itself, while simultaneously encouraging reuse and sharing of best practices.

4. CHALLENGES

We are currently using our PRISONER architecture for conducting a variety of SNS experiments involving sharing, privacy and context. Our system is far from complete, however. But we outline several outstanding challenges for further discussion.

1. *Is our architecture general enough for any SNS experiment?*

   We have designed our architecture to be extensible and to enable data collection from a variety of SNSs. It meets our experimental requirements, but it may not meet the requirements of all SNS researchers. We need to survey past experiments and survey researchers themselves to determine general requirements.

2. *How can we transform a policy into one that is understandable by non-researchers?*

   Our expectation is that researchers will design experimental policies and express them in XML. This might well be suitable for computer scientists who are experienced with such technologies. But can we assume that all researchers will be happy to do so? Can we build tools to enable policy design and what are their requirements? Can we use the XML policy to generate documentation (e.g., consent forms) that comply with the policy? Moreover, can we translate an XML policy into a human-readable policy that can be interpreted and reviewed by ethics committees?

3. *How can we sanitise collected SNS data?*

   Our experimental policies allow a researcher to express how data will be stored and sanitised. But it is well known that it is difficult to anonymise social network graph data (Narayanan & Shmatikov 2009). While anonymisation techniques are out of scope of our research, perhaps we need to build tools to visualise or describe the limits of various anonymisation and sanitisation strategies, both for researchers and participants.

4. *How can we respect participant requirements?*

   Research ethics typically revolve around the desires of research participants and mechanisms such as informed consent are designed to respect these desires. But it is well-known that people's privacy concerns can change over time, such as Barnes' "privacy paradox" (Barnes 2006), where SNS users become more concerned after a privacy breach. Similarly, Nissenbaum's analysis of privacy in terms of "contextual integrity" (Nissenbaum 2004) means that people might have different concerns for the same set of SNS data, depending on the context in which they are used (Hull et al. 2011). If concerns are variable, then should we allow participants to remove their data if they wish to withdraw consent? Can we do this with sanitised data?

5. *How can we encourage use of the system?*

   Perhaps the biggest challenge is convincing researchers that it is in their interests to use a system such as this. Some researchers have debated whether SNS research requires ethics approval at all (Solberg 2010), while other proposals for collecting SNS data have chosen to ignore privacy concerns (Cormode et al. 2010). Scientists may be loathe to consider ethics (Wolpe 2006), or perceive that they lack the time to go through the ethics approval process (Garfinkel & Cranor 2010). Our hope is that sharing of experimental policies may lead to development of best practices, which might well expedite the approval process, in addition to addressing other current concerns, such as helping committees to develop guidelines (Buchanan & Ess 2009) and educating researchers (Buchanan et al. 2011).

   Even if researchers do not wish to consider ethics, there may be legal compulsions; for instance the EU Data Protection Directive considers many SNS data to be *"sensitive"* and therefore subject to additional safeguards and processing (Edwards & Brown 2009). There may also be pressure from publication venues; for instance the 2012 Symposium on Usable Security and Privacy call for papers states

> "Authors may be asked to include explanation of how ethical principles were followed in their final papers should questions arise during the review process."[2]

5. CONCLUSION

Research using social network sites introduces a number of ethical concerns, and it can be hard for researchers and participants to consider all of these. We are developing an architecture for privacy-sensitive experiments. We hope that this can enable the collection and development of best practices, but in this paper we have outlined some challenges that remain.

6. REFERENCES

Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M., & Chalmers, M. (2008). From awareness to repartee: sharing location within social groups. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 497–506). doi:10.1145/1357054.1357134.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, *11*.

Buchanan, E., Aycock, J., Dexter, S., Dittrich, D., & Hvizdak, E. (2011). Computer science security research and human subjects: Emerging considerations for research ethics boards. *Journal of Empirical Research on Human Research Ethics*, *6*, 71–83. doi:10.1525/jer.2011.6.2.71.

Buchanan, E. A., & Ess, C. M. (2009). Internet research ethics and the institutional review board: current practices and issues. *ACM SIGCAS Computers and Society*, *39*, 43–49. doi:10.1145/1713066.1713069.

Cavoukian, A. (2011). Patience, persistence, and faith: Evolving the gold standard in privacy and data protection. In J. Camenisch, S. Fischer-Hübner, Y. Murayama, A. Portmann, & C. Rieder (Eds.), *Future Challenges in Security and Privacy for Academia and Industry* (pp. 1–16). Berlin, Germany: Springer. doi:10.1007/978-3-642-21424-0_1.

Cormode, G., Krishnamurthy, B., & Willinger, W. (2010). A manifesto for modeling and measurement in social media. *First Monday*, *15*.

Edwards, L., & Brown, I. (2009). Data control and social networking: Irreconcilable ideas? In A. M. Matwyshyn (Ed.), *Harboring Data: Information Security, Law, and the Corporation* chapter 10. (pp. 202–227). Palo Alto, CA, USA: Stanford University Press.

Garfinkel, S. L., & Cranor, L. F. (2010). Institutional review boards and your research. *Communications of the ACM*, *53*, 38–40. doi:10.1145/1743546.1743563.

Hull, G., Lipford, H. R., & Latulipe, C. (2011). Contextual gaps: privacy issues on Facebook. *Ethics and Information Technology*, *13*, 289–302. doi:10.1007/s10676-010-9224-8.

Hutton, L., & Henderson, T. (2012). An architecture for ethical and privacy-sensitive social network experiments. In *Proceedings of the ACM SIGMETRICS Workshop on Privacy and Anonymity for the Digital Economy*.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*, 330–342. doi:10.1016/j.socnet.2008.07.002.

Moreno, M. A., Fost, N. C., & Christakis, D. A. (2008). Research ethics in the MySpace era. *Pediatrics*, *121*, 157–161. doi:10.1542/peds.2007-3015.

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 173–187). doi:10.1109/SP.2009.22.

Neuhaus, F., & Webmoor, T. (2012). Agile ethics for massified research and visualization. *Information, Communication & Society*, *15*, 43–65. doi:10.1080/1369118X.2011.616519.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*, 101–139.

Office of the Information & Privacy Commissioner of Ontario, & Children's Hospital of Eastern Ontario (2011). Safeguarding personal health information when using mobile devices for research purposes.

Solberg, L. (2010). Data mining on Facebook: A free space for researchers or an IRB nightmare? *University of Illinois Journal of Law, Technology & Policy*, *2010*.

Toch, E., Cranshaw, J., Drielsma, P. H., Tsai, J. Y., Kelley, P. G., Springfield, J., Cranor, L., Hong, J., & Sadeh, N. (2010). Empirical models of privacy in location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 129–138). doi:10.1145/1864349.1864364.

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, *7*, 203–220. doi:10.1177/1745691612442904.

Wolpe, P. R. (2006). Reasons scientists avoid thinking about ethics. *Cell*, *125*, 1023–1025. doi:10.1016/j.cell.2006.06.001.

Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology*, *12*, 313–325. doi:10.1007/s10676-010-9227-5.

[2]http://cups.cs.cmu.edu/soups/2012/cfp.html