Usable data-driven privacy research: should we be afraid of the Big (Bad?) Data Wolf?

Tristan Henderson
School of Computer Science, University of St Andrews
St Andrews, Fife KY16 9SX
tristan@cs.st-andrews.ac.uk

1. INTRODUCTION

People in the "Surveillance Society" are surrounded by "Big Data". Privacy becomes harder and harder to protect as more and more data than ever before are collected about people's communications, mobility, social interactions and more. And yet, for many research questions, we still lack high-quality data. How, for instance, can we collect high-quality longitudinal data about users' privacy conceptions without affecting their behaviour, or without violating data protection laws? If we are to build and test usable systems, then the users of these systems must be able to trust that we, as researchers, have built systems that meet their ethical, legal and cultural expectations. Are there differences in behaviour, laws and culture that may require a European perspective that might be better met by a EuroSOUPS event?

2. BIG DATA: THE APPROPRIATE DATA?

There now exist several large datasets about behaviour and so forth, with websites such as infochimps.org and my own CRAWDAD network data archive[1] specialising in making such data available. Unfortunately, many of these datasets may not be the *right* data for some privacy research. As danah boyd puts it, "bigger data are not always better data" [6]. For instance, the ability to scrape large numbers of social network site profiles [14] might be useful for answering questions about privacy preferences, but it cannot tell us anything about the privacy behaviours that lead to data being withheld from an online social network, since by definition we can only gather those data that are shared. Thus work is needed in collecting high-quality, and perhaps longitudinal, data for understanding privacy, and thus enabling the design of usable privacy technologies.

3. BIG DATA: THE APPROPRIATELY COLLECTED DATA?

The ethical and legal implications of much research involving large datasets are still unclear. Some researchers have studied the EU Data Protection Directive in relation to social networks [7, 18] but the implications for research still remain unclear. Various high-profile studies [9] use data that have been collected in controversial ways, with some even being used as ethics case studies [8], or receiving the inevitable bad press [5, 15]. Thus work is needed to better understand how to design and conduct ethical and legal privacy studies.

4. PROPOSAL: BEST PRACTICES FOR DATA-DRIVEN PRIVACY RESEARCH

I therefore propose, and volunteer to contribute to, two goals for the development of the EuroSOUPS workshop:

1. The development of rigorous privacy experimental methodologies. I have argued above that we need to collect high-quality data. This is just one part of conducting rigorous, controlled and repeatable experiments. It would be good if EuroSOUPS could encourage studies which were rigorous in this regard. By helping to encourage and design suitable methodologies for controlled experiments, we may be able to build communities and infrastructures for repeating experiments and studies across multiple European sites.

2. The study of privacy within a European Data Protection environment. This is two-sided: it would be good to encourage awareness of the legal and ethical implications of the DPA, and also to study privacy concerns and behaviours from a European perspective, which may well differ from those in the US. This might even take the form of public policy involvement, as per SIGCHI.[2]

5. RESEARCH RECORD

My collaborators and I have extensive experience in measuring both networks [12] and users [10] and have been using this expertise to conduct user studies and collect high-quality data about privacy behaviours and concerns [2, 1, 3], as well as for developing new privacy-enhancing technologies such as user interfaces [13] and mobile routing schemes [17]. We have started exploring some of the different cultural and ethical issues of research such as data protection and attitudes to privacy but this is still preliminary [11], and as

[1] http://crawdad.org/

[2] http://www.sigchi.org/about/sigchi-public-policy

discussed above is something that I would like to explore further with a wider community.

Prior to my current position I was a Research Assistant Professor at Dartmouth College in the US and thus have experience of conducting studies and collecting data in both the US and European contexts. Relevant funded projects include *Privacy Value Networks* (EPSRC/TSB), *Digital Living: Sensors, Privacy and Trust* (US Department of Justice), *MAP (Measure, Analyze and Protect): Security through Measurement for Wireless LANs* (US Homeland Security Advanced Research Projects Agency).

In conjunction with Mike Just of Glasgow Caledonian University, I recently chaired a privacy and usability workshop, PUMP[3], at the BCS HCI conference. The attendees at this event expressed interest in EuroSOUPS and might be another useful starting point for building a EuroSOUPS community.

6. REFERENCES

[1] D. Anthony, T. Henderson, and J. Kitts. Trust and privacy in distributed work groups. In *Proc. 2nd Int'l Workshop on Social Computing and Behavioral Modeling*, pages 16–23, Phoenix, AZ, USA, Mar. 2009. DOI 10.1007/978-1-4419-0056-2_4.

[2] D. Anthony, T. Henderson, and D. Kotz. Privacy in location-aware computing environments. *IEEE Pervasive Computing*, 6(4):64–72, Oct. 2007. DOI 10.1109/MPRV.2007.83.

[3] F. Ben Abdesslem, I. Parris, and T. Henderson. Mobile experience sampling: Reaching the parts of Facebook other methods cannot reach. In *Proc. Privacy and Usability Methods Pow-Wow*, Dundee, UK, Sept. 2010. Online at http://scone.cs.st-andrews.ac.uk/pump2010/papers/benabdesslem.pdf.

[4] F. Ben Abdesslem, A. Phillips, and T. Henderson. Less is more: energy-efficient mobile sensing with SenseLess. In *Proc. ACM MobiHeld '09*, pages 61–62, Barcelona, Spain, Aug. 2009. DOI 10.1145/1592606.1592621.

[5] S. Borenstein. Cell phone users outside U.S. secretly tracked, June 2008. Associated Press. Retrieved 22/09/10 from http://www.msnbc.msn.com/id/24969880/.

[6] D. Boyd. Privacy and publicity in the context of big data. Talk at WWW '10: the 19th international conference on World wide web, Apr. 2010. Online at http://www.danah.org/papers/talks/2010/WWW2010.html.

[7] L. Edwards and I. Brown. Data control and social networking: Irreconcilable ideas? In A. M. Matwyshyn, editor, *Harboring Data: Information Security, Law, and the Corporation*. Stanford University Press, Palo Alto, CA, USA, 2009. Online at http://ssrn.com/abstract=1148732.

[8] GISProfessional Ethics Project. Case study: Tracking mobile phones in mobility research, Nov. 2009. Retrieved 22/09/10 from https://www.e-education.psu.edu/files/sites/file/mobile_phone_tracking_case%281%29.pdf,.

[9] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008. DOI 10.1038/nature06958.

[10] T. Henderson, D. Anthony, and D. Kotz. Measuring wireless network usage with the experience sampling method. In *Proc. WiNMee*, Trentino, Italy, Apr. 2005. International Communications Sciences and Technology Association (ICST). Online at http://www.winmee.org/2005/papers//WiNMee_Henderson.pdf.

[11] T. Henderson and F. Ben Abdesslem. Scaling measurement experiments to planet-scale: ethical, regulatory and cultural considerations. In *Proc. ACM HotPlanet '09*, pages 1–5, Kraków, Poland, June 2009. DOI 10.1145/1651428.1651436.

[12] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712, Oct. 2008. DOI 10.1016/j.comnet.2008.05.003.

[13] A. Kapadia, T. Henderson, J. Fielding, and D. Kotz. Virtual walls: Protecting digital privacy in pervasive environments. In *Proc. Pervasive 2007*, pages 162–179, Toronto, Canada, May 2007. DOI 10.1007/978-3-540-72037-9_10.

[14] K. Lewis, J. Kaufman, and N. Christakis. The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, 14(1):79–100, Oct. 2008. DOI 10.1111/j.1083-6101.2008.01432.x.

[15] P. Lewis. Bluetooth is watching: secret study gives Bath a flavour of Big Brother. *The Guardian*, July 21, 2008. http://www.guardian.co.uk/uk/2008/jul/21/civilliberties.privacy.

[16] I. Parris, F. Ben Abdesslem, and T. Henderson. Facebook or Fakebook?: The effect of simulation on location privacy user studies. In *Proc. Privacy and Usability Methods Pow-Wow*, Dundee, UK, Sept. 2010. Online at http://scone.cs.st-andrews.ac.uk/pump2010/papers/parris.pdf.

[17] I. Parris and T. Henderson. Privacy-enhanced social-network routing. *Computer Communications*, June 2010. Accepted for publication.

[18] B. Van Alsenoy, J. Ballet, A. Kuczerawy, and J. Dumortier. Social networks and web 2.0: are users also bound by data protection regulations? *Identity in the Information Society*, 2(1):65–79, Dec. 2009. DOI 10.1007/s12394-009-0017-3.

[3] http://scone.cs.st-andrews.ac.uk/pump2010/