Research Ethics in the Big Data Era: Addressing Conceptual Gaps for Researchers and IRBs

Michael Zimmer, PhD
School of Information Studies
University of Wisconsin – Milwaukee
zimmerm@uwm.edu

Introduction

We have entered the era of big data. We can now access petabytes of transaction data, clickstreams and cookie logs, media files, and digital archives, as well as data from social networks, mobile phones, and sensors. This data is growing exponentially, as is the technology to extract insights, discoveries, and meaning from it. Big data promises to transform every sector of the economy. Cloud storage, fast networks, and advanced analytic software now offer professionals access to real-time data. Data-driven decision-making has already increased productivity in financial services and social media. Health care, public administration, retail, and manufacturing are poised to achieve similar growth as they harness the technologies underlying this data revolution. To date, computer scientists, mathematicians, and statisticians working in finance and information industries have dominated this new data science. But the tools for working with big data are improving fast, and a much wider range of sectors, including health care, manufacturing, education and government, are now in pursuit of the value of data-driven decision making that big data promises. Further, big data has emerged as a rich terrain for engaging in research and experimentation – both in academic and commercial environments – yielding novel results, while also generating considerable controversy.

In his foundational essay, "What is Computer Ethics?," James Moor (1985) notes how the malleable nature of computer technology – the ease at which it can be shaped and molded for use in a variety of unexpected ways – will transform "many of our human activities and social institutions," and will "leave us with policy and conceptual vacuums about how to use computer technology" (p. 272). As a

result, Moor argues, we are left with little guidance on how to address the new ethical dilemmas that inevitably arise with the increased use of computer technology:

> A typical problem in Computer Ethics arises because there is a policy vacuum about how computer technology should be used. Computers provide us with new capabilities and these in turn give us new choices for action. Often, either no policies for conduct in these situations exist or existing policies seem inadequate. A central task of Computer Ethics is to determine what we should do in such cases, that is, formulate policies to guide our actions. (Moor, 1985, p. 266)

Today, we are confronted with innumerable policy vacuums and conceptual gaps triggered by big data's astonishing ability to transform "many of our human activities and social institutions," far beyond what Moor envisioned with computing technology 30 years ago.

Attempts to fill the policy vacuums and clarify the conceptual gaps created in the wake of big data have already begun, stemming largely from concerned computer scientists, social scientists, and legal scholars. This essay helps push forward the growing discourse on the ethics of big data research by disclosing critical conceptual gaps that often hamper how researchers and IRBs think about privacy, personal information, consent, and harm in the context of big data. In doing so, this essay follows a disclosive approach to computer ethics research (Brey, 2000), uncovering ethical issues in big data analytics and research that have not yet gained sufficient attention, and making visible the conceptual gaps that have emerged. Through such attempts to address and clarify these conceptual gaps we can strive to ensure that ethical guidelines and policies are properly formulated to support the central concerns of research ethics and human subject protection amid the rise of big data research.

Conceptual Gaps: Privacy, Personally Identifiable Information, Consent, and Harm

Privacy is typically protected within the context of human subjects research through a combination of various tactics and practices, including engaging in data collection under controlled or anonymous environments, limiting the personal information gathered, scrubbing data to remove or obscure personally identifiable information, and using access restrictions and related data security methods to prevent unauthorized access and use of the research data itself. The nature and understanding

of privacy becomes muddled, however, in the context of big data research, and as a result, ensuring it is respected and protected in this new domain becomes challenging.

For example, the determination of what constitutes "private information"—and thus triggering particular privacy concerns—becomes difficult within the context of big data research. Distinctions within the regulatory definition of "private information" – that it only applies to information which subjects reasonably expect is not normally monitored or collected and not normally publicly available – become less clearly applicable when considering the data environments and collection practices that typify big data research, such as the wholesale mining of Facebook activity or public Twitter streams. When considered through the lens of the regulatory definition of "private information," social media postings are typically considered public, especially when users take no steps to restrict access, and are thus not deserving of particular privacy consideration. For example, researchers in the Harvard "Tastes, Ties, and Time" research project (Lewis, et al, 2008) – where an entire cohort of college students had their Facebook profiles scraped annually for for years – argued that subjects do not have a reasonable expectation of privacy with their Facebook information, noting "We have not accessed any information not otherwise available on Facebook," and equating their collecting of the profile data with "sitting in a public square, observing individuals and taking notes on their behavior" (comment at Zimmer, 2008b). Similarly, much of the discussion surrounding the appropriateness of harvesting Twitter activity centers on the basic fact that a public Twitter stream is purposefully visible to anyone, thus no privacy expectations exist (see, for example, discussions at 2008b).

Yet, the social media platforms frequently used for big data research purposes represent a complex environment of social interaction where users are often required to place friends, lovers, colleagues, and minor acquaintances within the same singular category of "friends," where privacy policies and terms of service are not fully understood (Madejski, Johnson, & Bellovin, 2011), and where the technical infrastructures fail to truly support privacy projections (Bonneau & Preibusch, 2009) and regularly change with little notice (Stone, 2009; Zimmer, 2009). Similarly, numerous studies have indicated that average Internet users have incomplete understandings of how their activities are routinely

tracked and the related privacy practices and policies of the sites they visit (Hoofnagle & King, 2008; Milne & Culnan, 2004; Tsai, Cranor, Acquisti, & Fong, 2006). As a result, it is difficult to understand with certainty what a user's intention was when posting an item on a social media platform (Acquisti & Gross, 2006). Thus, it remains unclear whether Internet users truly understand if and when their online activity is regularly monitored and tracked, and what kind of reasonable expectations truly exist. This uncertainty in the intent and expectations of users of social media and internet-based platforms—often fueled by the design of the platforms themselves—create a conceptual gap in our ability to apply the definition of "private information" to ensure subject privacy is properly addressed and forces us to reconsider the justification "we have not accessed any information not otherwise available" in order to alleviate potential privacy concerns.

Similar conceptual gaps emerge when we consider the traditional definitions of "personally identifiable information" in the context of big data research, where there are increased pressures to release datasets, as well as increased opportunities to access and combine databases from various sources. Increasingly, datasets considered "anonymized" have been re-identified, often with relative ease, relying on information not covered under the regulatory definition of "personally identifiable." For example, researchers have been able to re-identify individuals by analyzing and comparing such datasets, using data-fields as benign as one's ZIP code (Sweeney, 2002), random Web search queries (Barbaro & Zeller Jr, 2006), or movie ratings (Narayanan & Shmatikov, 2009) as the vital key for reidentification of a presumed anonymous user. Prior to widespread Internet-based data collection and processing, few would have considered one's movie ratings or ZIP code as personally identifiable. Yet, merely stripping traditional "identifiable" information such as a subject's name, address, or social security number is no longer sufficient to ensure that data remains anonymous (Ohm, 2009) and require the reconsideration of what is considered "personally identifiable information" (Schwartz & Solove, 2011).

The conceptual gaps that exist regarding privacy and the definition of personally identifiable information in the context of big data research inevitably lead to similar gaps regarding when informed consent is necessary. Researchers mining Facebook profile information or public Twitter streams, for

example, typically argue that no specific consent is necessary due to the fact the information was publicly available. Yet, it remains unknown whether users truly understood the technical conditions under which they made information visible on these social media platforms (see Hoofnagle and King 2008; Acquisti and Gross 2006) or if they foresaw their data being harvested for research purposes, rather than just appearing onscreen for fleeting glimpses by their friends and followers. In the case of the ill-fated Facebook emotional contagion experiment (Kramer, Guillory, & Hancock, 2014), the lack of obtaining consent was initially rationalized through the notion that the research appeared to have been carried out under Facebook's extensive terms of service, whose data use policy, while more than 9,000 words long, does make passing mention to "research." It was later revealed, however, that the data use policy in effect when the experiment was conducted never mentioned "research" at all (Hill, 2014).

The growing domain of big data research has also led to conceptual gaps in how we define what actually might be a harm and the human subjects themselves deserving protection. Extending from the common arguments again any persistent privacy concerns with accessing and sharing subject information gathered from various data sources that fuel big data analytics, researchers also frequently suggest harms are not present nor are imminently forthcoming when subject data is already readily available online for anyone to access and use.

Such positions equating harm to tangible loss or impact on a subject ignore the broader dignity-based theory of privacy harm (Bloustein, 1964). Such a stance recognizes that one does not need to be a victim of hacking or have a tangible harm take place in order for there to be concerns over the privacy of one's personal information. Rather, merely having one's personal information stripped from the intended sphere of the social networking profile and amassed into a database for external review becomes an affront to the subjects' human dignity and their ability to control the flow of their personal information. This conceptual gap is not unique to research ethics, as international laws and regulations surrounding the collection and use of personal data similarly vary as to the definition of harm (Bennett & Raab, 2006). For example, Canadian and European Union regulations embrace a largely paternalist approach to data protection policy, aiming to preserve a fundamental human right of its citizens through preemptive

governmental action, believing users must maintain control over their information to preserve dignity and autonomy. In contrast, the governance of privacy in the United States begins with the assumption that most data collection and use is both acceptable and beneficial, and limits are imposed only after some tangible harm has occurred.

Those embracing a more European approach to harm acknowledge that threats to a subject's dignity or autonomy are as meaningful as more tangible harms such as exposure of personal information or identity theft. But coming to this conclusion requires, at the start, the recognition that human subjects themselves are at risk within the design of Internet-based research studies. In various cases, opinions differ whether human subjects are even involved in Internet-based projects, providing a particularly potent conceptual gap for researches and IRBs to contend with. Much of the debate surrounding the ethics of archiving public Twitter streams centers on whether tweets are public utterances by human subjects, and thus requiring ethical review, or merely the equivalent of published texts, thus exempted from any ethical concern (see discussion at Zimmer 2010b). Similarly, researchers studying large datasets or communication network traffic, for example, frequently perceive their studies as outside the purview of IRBs, since, in their view, the IRB review process is "used more in medical and psychology research at our university" (as quoted in Soghoian, 2012) or perceive IRBs as bothersome barriers to achieving important research outcomes (Garfinkel, 2008).

To help address this fundamental conceptual gap, Carpenter and Dittrich (2011) introduce the notion of "human-harming research" as a variable in human subjects review in Internet-based and computer science research. They worry that researchers increasingly perceive an increased "distance" between themselves and their subjects; rather than researchers engaging with subjects directly, interactions and data-collection are increasingly mediated by social media profiles, data networks, and transaction logs. As a result, the perception of a human subject becomes diluted through increased technological mediation. To compensate, Carpenter and Dittrich (2011) encourage ethical review boards to transition "from an informed consent driven review to a risk analysis review that addresses potential harms stemming from research in which a researcher does not directly interact with the at-risk

individuals" (4), and to ultimately "transition our idea of research protection from 'human subjects research' to 'human harming research'" (14). In doing so, researchers who might otherwise (even if incorrectly) feel no human is directly involved in the research study would be compelled to address the ethical implications of any harm to broader populations outside the immediate research project.

Conclusion

The growing use of the Internet as both a research tool and a site for research itself has produced a deluge of unique and innovative big data-based research projects and methodologies which test the ethical frameworks and assumptions traditionally used by researchers and IRBs to ensure adequate protection of human subjects. The result is numerous conceptual gaps in how to apply established research ethics principles in the context of social media research. This essay sought to disclose some of the ethical concerns with big data research, making transparent some of the emerging conceptual gaps. The ethical, research and regulatory communities must engage in collaborative, dedicated, and multi-prong efforts focusing on four objectives: further disclosing the conceptual gaps present in big data research, reframing the ethical dilemmas inherent in such research projects, expanding educational and outreach efforts, and developing policy guidance focused on the unique challenges of big data research ethics. By attending to such concerns, researchers and IRBs will be better positioned to understand and address the ethical dimensions of big data research projects, close the existing conceptual gaps, and thereby ensure innovative research can take place while protecting the interests of research ethics broadly.

References

Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Proceedings of the 6th Workshop on Privacy Enhancing Technologies*, *4258*, 36–58.

Barbaro, M., & Zeller Jr, T. (2006). A Face Is Exposed for AOL Searcher No. 4417749. *The New York Times*, A1.

Bennett, C. J., & Raab, C. D. (2006). *The governance of privacy: Policy instruments in global perspective*. Cambridge, MA: MIT Press.

Bloustein, E. (1964). Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser. *New York University Law Review*, *39*, 962–1007.

Bonneau, J., & Preibusch, S. (2009). The Privacy Jungle: On the Market for Data Protection in Social Networks. *The Eighth Workshop on the Economics of Information Security (WEIS 2009)*.

Brey, P. (2000). Disclosive computer ethics. *Computers and Society*, *30*(4), 10–16.

Carpenter, K., & Dittrich, D. (2011). Bridging the Distance: Removing the Technology Buffer and Seeking Consistent Ethical Analysis in Computer Security Research. *1st International Digital Ethics Symposium*.

Garfinkel, S. (2008). IRBs and security research: Myths, facts and mission creep. *Proceedings of the 1st Conference on Usability, Psychology, and Security*. Retrieved from http://static.usenix.org/event/upsec08/tech/full_papers/garfinkel/garfinkel.pdf

Hill, K. (2014, June 30). Facebook Added "Research" To User Agreement 4 Months After Emotion Manipulation Study. Retrieved November 6, 2015, from http://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/

Hoofnagle, C. J., & King, J. (2008). What Californians understand about privacy online. *Research Report. Samuelson Law Technology & Public Policy Clinic UC Berkeley Law: Berkeley, CA*.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. http://doi.org/10.1073/pnas.1320040111

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*(4), 330–342.

Madejski, M., Johnson, M., & Bellovin, S. (2011). The failure of online social network privacy settings. *Posterouscom*, *CUCS-010-11*, 1–20.

Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, *18*(3), 15–29.

Moor, J. (1985). What is computer ethics? *Metaphilosophy*, *16*, 266–275.

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing Social Networks. *30th IEEE Symposium on Security and Privacy*.

Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701.

Schwartz, P., & Solove, D. (2011). The PII Problem: Privacy and a New Concept of Personally Identifiable Information. *New York University Law Review*, *86*, 1814.

Soghoian, C. (2012). Enforced community standards for research on users of the Tor anonymity network. *Financial Cryptography and Data Security*, *7126*, 146–153.

Stone, B. (2009). Facebook Rolls Out New Privacy Settings. Retrieved from http://bits.blogs.nytimes.com/2009/12/09/facebook-rolls-out-new-privacy-settings/

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570.

Tsai, J., Cranor, L., Acquisti, A., & Fong, C. (2006). What's it to you? a survey of online privacy concerns and risks. *NET Institute Working Paper No. 06-29*.

Zimmer, M. (2008a). More On the “Anonymity” of the Facebook Dataset - It’s Harvard College. Retrieved from http://michaelzimmer.org/2008/10/03/more-on-the-anonymity-of-the-facebook-dataset-its-harvard-college/

Zimmer, M. (2008b). On the “Anonymity” of the Facebook Dataset. Retrieved from http://michaelzimmer.org/2008/09/30/on-the-anonymity-of-the-facebook-dataset/

Zimmer, M. (2009). FacebookÕs Privacy Upgrade is a Downgrade for User Privacy. Retrieved from http://michaelzimmer.org/2009/12/10/facebooks-privacy-upgrade-is-a-downgrade-for-user-privacy/