2018

Protecting Human Subjects in the Digital Age: Issues and Best Practices of Data Protection

Thomas Jamieson
University of Waterloo, tjamieson@unomaha.edu

Güez Salinas
University of Southern California

Vol. 11, Issue 2, 2018

Protecting Human Subjects in the Digital Age: Issues and Best Practices of Data Protection

Thomas Jamieson[*], Güez Salinas[†]

[*] **Institution:** University of Waterloo **Department:** Department of Political Science
ORCID iD: 0000-0002-2716-5476
[†] **Institution:** University of Southern California (CA) **Department:** Political Science and International Relations

Abstract

Public opinion and survey researchers must protect the privacy and confidentiality of human subjects. However, scholars are often not trained in the best practices of data storage, and there is a serious risk that survey data might be compromised by pernicious actors. In an era when it is becoming increasingly difficult to recruit participants, breaches could further challenge our ability to conduct surveys if we cannot guarantee that participants' data will remain confidential and private. While any computer-based data has some vulnerability, we introduce simple measures that will better protect the confidentiality and privacy of human subjects. We hope these could become standard practice to protect human subjects in the future.

Introduction

Online tools and recruitment platforms have made data collection for survey and experimental research increasingly accessible and inexpensive. Similarly, cloud storage and improved technology make data storage more affordable and convenient than ever. However, advances in computer-based survey tools and methods come with challenges regarding the protection of human subjects' confidentiality and privacy (Conway and O'Connor 2016; Couper 2017; Hashem et al. 2015; Itani et al. 2009; Link et al. 2014). Participants are provided with information that scholars will remove identifiable information from their survey responses, and they proceed in the expectation that their identity will be protected. However, despite the nontrivial risks that persist, there is little guidance on the best practices for survey researchers on data security and storage.

Public opinion and survey researchers are not always up to date with the best practices of data storage and digital communication, but there is a serious risk of data being compromised by pernicious actors. Recent breaches include hacks of Democratic National Committee data, and there have been hundreds of successful attacks on various organizations across different sectors, exposing the data of billions of people (Information is Beautiful 2018). The threat of data breaches is significant for public opinion research, presenting fundamental challenges for researchers.

In an era when it is becoming increasingly difficult to recruit participants, breaches could further challenge public opinion scholars' ability to recruit participants and conduct research if we cannot ensure that participants' data remains confidential and private (Acquisti et al. 2015). At best, it opens researchers up to unnecessary liability when research is conducted.

This is particularly important given increased state and governmental acknowledgement about the rights of citizens to privacy and the responsibilities of institutions to protect personal data. For example, the newly-implemented European Union General Data Protection Regulation (GDPR) requires that personal data is "processed in a manner that ensures appropriate security of the personal data, including protection against

unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures" (European Parliament, Council of the European Union 2016: 35–36). In short, improved data protection practices are necessary given the convergence of increasing threats to data protection, increasing difficulty in recruiting participants for survey research, and increasing ethical and legal burdens on organizations to protect personal data.

In this in-brief note, we outline some issues with data protection while storing and transmitting in the digital age that we have seen working as both social scientists and network engineers. To address these threats, we then introduce simple prescriptive measures researchers can take to, at the minimum, improve the level of individual digital hygiene, and at best raise institutionalized standard operating procedures to mitigate against existing threats. To supplement these initial measures, we conclude the note with a list of resources for further reading and reflection on data protection.

Issues with Data Protection

Threats to Privacy

Privacy refers to participants' ability to control how their information is collected and how it is used. Threats to privacy could include the collection of data without participants' explicit and informed consent, or breaches where collected data becomes accessible to people not identified at the time of collection. It is vital that researchers ensure the privacy of participants, ensuring that any data collected is used according to the specific purpose identified in recruitment materials.

Researchers can take standard precautions such as ensuring participants are informed about the survey and storing data on password-protected computers on secure servers. However, if data can be accessed via the Internet, or through an open Application Programming Interface (API), it is susceptible to being accessed from pernicious actors (Liu et al. 2012). This is particularly concerning as researchers adopt cloud-based storage (Chen and Zhao 2012; Itani et al. 2009), work within facilities

that have attracted several high-profile attacks that have affected millions of users (Reilly 2014; BBC News 2014), and work in institutions that have been targets for hackers (Kopan 2018). Specific threats to data privacy in survey research include access for colleagues in the same cloud environment, human error revealing private information, glitches in connectivity, or external hacks (Barona and Anita 2017). Scholars face important challenges to ensure the privacy of data given these vulnerabilities.

Threats to Confidentiality

A related concern is the confidentiality of data. Confidentiality refers to the risk that participants' identity might be revealed, especially if this is counter to the conditions under which they participated in the survey. Even if scholars de-identify individual-level data, if malicious actors access it they might be able to re-identify participants, especially if the coding scheme is present (Clauß et al. 2005). As Narayanan and Shmatikov (2010, 26) warned, "Advances in the art and science of reidentification, increasing economic incentives for potential attackers, and ready availability of personal information about millions of people (for example, in online social networks) are rapidly rendering [de-identification] obsolete." As such, even if researchers adopt current best practices (AAPOR 2018), participants' confidentiality is at risk.

While survey researchers may not be able to eliminate risk entirely, researchers can take precautions to mitigate against these threats as best as possible. We have to keep in mind that given the nature of how computers and other networked devices communicate with one another on the Internet, the complete elimination of all vulnerabilities is arguably impossible. However, we can take precautions that make the likelihood of a vulnerability being nefariously exploited by a maliciously motivated actor significantly more difficult. The point is to make your data a less inviting target.

Best Practices of Data Protection

Ultimately, any data on computers that are connected to the Internet possess some vulnerability to breaches of privacy and/

or confidentiality for participants' data. In this section, we recommend several simple measures that will help address these critical threats to data privacy and confidentiality, decreasing the likelihood of a levied attack being successful. These practices supplement existing research guidelines provided by the American Association for Public Opinion Research, providing greater detail about protecting data from potential breaches. We break these measures down by the stage of the survey research process:

Data Collection

- Encrypt all communication with participants. This is particularly important when communicating via email, which is especially vulnerable to breaches from external actors and social engineering attacks. Given the standard heuristics involved in all malicious hacks regardless of kind, seemingly innocuous email communication has the potential to give an attacker useful information once the desired data is stolen. Simple email add-ons such as PGP Encryption, an open-source email encryption tool that make sure emails are unreadable without having a decryption key, will render any hacked communications undecipherable without the proper decryption keys.

- Assign ID numbers to participants separately.

Data Storage

- Store all data using encrypted, password-protected storage services and/or drives (e.g., File Vault or its equivalent).

- If possible, avoid storing data in the cloud.

- At the very least, any codebooks or documents containing data with identifiers should be stored separately on encrypted storage devices or services. Sampling frames and response datasets should also be stored separately. This complicates the deanonymization process for a hacker that has gained unauthorized access, where they would need to breach several firewalls to

gain access to personal data. Ideally, codebooks with identifiers should be stored physically and not electronically at all.

- Use separate computers for work and personal usage, or at the very least partition drives to increase the security of your device. Doing so will limit the level of compromise if one of the machines is actually compromised, in essence quarantining the intrusion.

Conclusion

Public opinion scholars face new challenges to ensure human subjects' data remains private and confidential in the hyper-connected digital age. Fortunately, a few simple measures can greatly improve data security for survey researchers, helping to keep participants' data private and confidential. We hope these could become standard practice to protect human subjects' privacy and confidentiality in the future. However, as new technology becomes adopted and new threats emerge, the best practices will change, so it is important for survey researchers to stay engaged in issues of data protection.

The Internet, at its core, is the same as the original ARPANET from the late 1950s. We need to raise the awareness of just how inherently vulnerable the overall architecture is and take collective action to move toward encryption behaviors that make data, even if stolen, impossible to read.

Further Reading

For further reading about threats to data protection, and how to mitigate against the risk of data breaches, please refer to the following guides and resources:

Guidelines for Research

- American Association for Public Opinion Research [AAPOR]. 2018. Best practices for survey research. Available at https://www.aapor.org/Standards-Ethics/Best-Practices.aspx.

- European Parliament, Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance). *Official Journal of the European Union 119* (4.5.2016): 1–88 Available at https://eur-lex.europa.eu/eli/reg/2016/679/oj

U.S. Federal Guidelines

- National Institute of Standards and Technology. 2004. FIPS 199, Standards for security categorization of federal information and information systems. Retrieved July 29th, 2018 from https://csrc.nist.gov/publications/detail/fips/199/final

- National Institute of Standards and Technology. 2006. FIPS 200, Minimum security requirements for federal information and information systems. Retrieved July 29th, 2018 https://csrc.nist.gov/publications/detail/fips/200/final.

- U.S. Department of Homeland Security. 2015. US-CERT Federal Incident Notification Guidelines | US-CERT. Retrieved July 29th, 2018. https://www.us-cert.gov/incident-notification-guidelines

- U.S. Department of Homeland Security. 2016. Automated Indicator Sharing (AIS) | US-CERT. Retrieved July 29th, 2018. https://www.us-cert.gov/ais

Recommended Further Training

- International Academic Program $(ISC)^2$. 2018. International Academic Program. Retrieved July 29th, 2018. https://www.isc2.org:443/IAP

Important Questions to Ask Your IT Team

- Whether maintaining your own servers or using a cloud

service, has your systems administrator changed default passwords to sufficiently complex passwords?

- What is the level of encryption of all stored data?
- Is all information communication traffic being sent through secure *https* connections, as opposed to *http*?
- Does your organization store sensitive data on personal use computers that can be easily compromised through social engineering based attacks?

References

Acquisti, A., L. Brandimarte, and G. Loewenstein. 2015. "Privacy and Human Behavior in the Age of Information." *Science* 347 (6221): 509–514.

American Association for Public Opinion Research [AAPOR]. 2018. "Best Practices for Survey Research." https://www.aapor.org/Standards-Ethics/Best-Practices.aspx.

Barona, R., and E.A.M. Anita. 2017. "A Survey on Data Breach Challenges in Cloud Computing Security: Issues and Threats." In *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT*, 1–8. https://doi.org/10.1109/ICCPCT.2017.8074287.

BBC News. 2014. "FBI Probes 'Cloud' Celebrity Leaks," September 2, 2014. http://www.bbc.com/news/technology-29011850.

Chen, D., and H. Zhao. 2012. "Data Security and Privacy Protection Issues in Cloud Computing." In *Proceedings of the 2012 International Conference on Computer Science and Electronics Engineering*, 01:647–651. Washington, DC: IEEE Computer Society.

Clauß, S., D. Kesdogan, and T. Kölsch. 2005. "Privacy Enhancing Identity Management: Protection against Re-Identification and Profiling." In *Proceedings of the 2005 Workshop on Digital Identity Management, DIM '05*, 84–93. New York: ACM.

Conway, M., and D. O'Connor. 2016. "Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications." *Current Opinion in Psychology* 9 (June): 77–82.

Couper, M.P. 2017. "New Developments in Survey Data Collection." *Annual Review of Sociology* 43 (1): 121–145.

European Parliament, Council of the European Union. 2016. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA

Relevance." *Official Journal of the European Union* 119 (4.5.2016): 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Hashem, I.A.T., I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, and S.U.l. Khan. 2015. "The Rise of 'Big Data' on Cloud Computing: Review and Open Research Issues." *Information Systems* 47 (January): 98–115.

Information is Beautiful. 2018. "Data Breaches (Public)." Google Docs. 2018. https://docs.google.com/spreadsheets/d/1Je-YUdnhjQJO_13r8iTeRxpU2pBKuV6RVRHoYCgiMfg/edit?usp=embed_facebook.

Itani, W., A. Kayssi, and A. Chehab. 2009. "Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures." In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, 711–716. Washington, DC: IEEE Computer Society.

Kopan, T. 2018. "Iranians Hacked Thousands of US Professors, Justice Dept. Says - CNNPolitics." CNN. 2018. https://www.cnn.com/2018/03/23/politics/iranian-hackers-indicted-universities-government/index.html.

Link, M.W., J. Murphy, M.F. Schober, T.D. Buskirk, J.H. Childs, and C.L. Tesfaye. 2014. "Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research." *Public Opinion Quarterly* 78 (4): 779–787.

Liu, J., Y. Xiao, S. Li, W. Liang, and C.L.P. Chen. 2012. "Cyber Security and Privacy Issues in Smart Grids." *IEEE Communications Surveys Tutorials* 14 (4): 981–997. doi: 10.1109/SURV.2011.122111.00145.

Narayanan, A., and V. Shmatikov. 2010. "Myths and Fallacies of 'personally Identifiable Information." *Communications of the ACM* 53 (6): 24–26.

Reilly, C. 2014. "Hackers Hold 7 Million Dropbox Passwords Ransom." 2014. https://www.cnet.com/news/hackers-hold-7-million-dropbox-passwords-ransom/.