Assignment 8

MACS 30000

Li Liu

11/27/2018

1. Identification risk in anonymized data

    (a) Sweeney (2012) shows that sensitive medical records information could be re-identified by combining the "anonymized" health records and public voting records. Zimmer (2010) points out the identities in the "Tastes, Ties, and Time" project data are students at Harvard College. In both cases, the re-identification processes were achieved by matching the unique characteristics (such as zip code, birthday, sex, ethnicity, race) in the original dataset with other available data containing this information.

    (b) Sweeney (2012) illustrate the danger of public healthcare information by identifying William Weld (68th Governor of Massachusetts)'s personal medical records. Zimmer (2010, p. 314) argues that the re-identified data has sensitive information including housing, demographic, cultural, relational information and even names for individuals. People's privacy and safety are put in risk if this information is maliciously used by some organizations and hackers.

2. Describing ethical thinking

Kauffman (Sep. 30, 2008b), the principal investigator of the "Tastes, Ties, and Time" research project, justified for good ethics of publishing their data by stating they are "sociologists, not technologists". However, the problem is about ethics instead of technology. He pointed out that sociologists are interested in studying people (Kauffman, Sep.30, 2008b). This violates the principle of "Respect for Persons" as the research subjects might not consent to be studied (Salganik, 2018, p. 295). Kauffman was following the framework of "consequentialism" (Salganik, 2018, p. 302) as he thought the data would be "enormously useful to researchers interested in taste, culture, etc." (Sep. 30, 2008b). He focused on the expected outcome of researches instead of the potential risk of revealed sensitive information. He also defended the validity of their aggregated dataset as it "contains almost no information that isn't on Facebook" (Kauffman, Sep.30, 2008b). This impairs the "justice" principle (Salganik, 2018, p. 298) as students at Harvard College would bear more risks than other students.

In the second comment, Kauffman (Sep. 30, 2008c) was content with fulfilling IRB's rules. He considered IRB as "ceiling" instead of the "floor" and didn't take enough ethnical responsibility (Salganik, 2018, p. 322). He iterated that researches didn't "have to ask those individuals' consent in advance" since they didn't "make information about them public" (Kauffman, Sep.30, 2008c). His argument didn't show essential "respect for law and public interest" (Salganik, 2018, p. 299). He didn't consider the terms-of-service agreement by Facebook, nor did he value individuals' privacy. Even though researchers could find insightful knowledge from the data, they should protect the research objects from malicious attacks.

3. Ethics of Encore

(a) Narayanan and Zevenbergen (2015) recognize Encore study (Burnett and Feamster, 2015)'s significance in innovating data collection techniques and studying the Internet censorship of governments, but criticize the study for changing browsing behaviors and lacking users' consent. Ethical issues would become more important as researchers are getting interested in individuals and organizations' behaviors from analyzing Internet network data and measuring censorship (Narayanan and Zevenbergen, 2015). They argue that the program reviewing committees should pay more attention to deal with projects that humans and technology interact (Narayanan and Zevenbergen, 2015). Narayanan and Zevenbergen (2015, p.12) adopt the "consequentialist" framework (Salganik, 2018, p. 302) in their ethical assessment of the Encore project as they categorize Encore's data is about "behavior of censorship systems" instead of "the individual". They also apply the "principle of beneficence" by evaluating the risks and benefits of the Encore project and suggest researchers mitigate as much harm as possible. (Narayanan and Zevenbergen, 2015, p. 13-19).

(b) Burnett and Feamster (2015)'s Encore study calls for more attention in research ethics. I am very doubtful about the good ethics of this project. If my browser is forced to connect to some blocked websites, I feel being unethical and in danger for potential consequences. In addition, some blocked websites would be risky and users' privacy information might be leaked by malicious viruses.

Reference

Burnett, Sam and Nick Feamster, "Encore: Lightweight Measurement of Web Censorship with

   Cross-Origin Requests," 2015.

Kauffman, Jason, "I am the Principle Investigator...," Blog Comment, MichaelZimmer.org,

   http://www.michaelzimmer.org/2008/09/30/ on-the-anonymity-of-the-facebook-dataset/, Sep.

   30, 2008b.

---, "We did not consult...," Blog Comment,

   MichaelZimmer.org,http://www.michaelzimmer.org/2008/09/30/

   on-the-anonymity-of-the-facebook-dataset/, Sep. 30, 2008c.

Montjoye, Yves-Alexandre de, Laura Radaelli, Vivek Kumar Singh, and Narayanan,

   Arvind and Bendert Zevenbergen, "No Encore for Encore? Ethical Questions for

   Web-based Censorship Measurement," Technology Science, December 15 2015.

Salganik, Matthew J., Bit by Bit: Social Research in the Digital Age, Princeton University

   Press, 2018.

Sweeney, Latanya, "K-Anonymity: A Model for Protecting Privacy," International Journal on

   Uncertainty Fuziness and Knowledge-Based Systems, 2002, 10 (5), 557– 570.

Zimmer, Michael, "But the Data is Already Public: On the Ethics of Research in Facebook,"

   Ethics and Information Technology, 2010, 12 (4), 313–325.