Assignment #8

MACS 30000, Dr. Evans

Due Monday, Dec. 3 at 11:30am

Ruixi Li

1. Identification risk in anonymized data

a) I choose Health insurance records from Sweeney (2002) and Demographic, administrative, and social data about students from Zimmer (2010). In both cases, the re-identification attach has a similar structure. That is matching anonymized data with identified data by some shared variables. In Sweeney (2002), by combining the unnamed health records and public voting records, it re-identified insurance records. In Zimmer (2010, p314) shows that the samples in the anonymized project data are Harvard College students.

b) In Sweeney (2002), the dataset has medical information about anonymized individuals like the medical visit date, diagnosis, procedure, medication, and total medical charge and some non-identifying variables like ZIP code, sex, and date of birth. Although the name of the individual is sealed, with an additional dataset of voter registration data which contain some shared attributes with medical dataset and individual's sensitive information, the writer can merge this two dataset and re-identify the individual. This kind of identification may induce leakage of health information.

In Zimmer (2010), based on the incident of realizing anonymized profile data, the author argues that although the identifying information was hidden. It may be re-identified and retrieved. If so, the sensitive information like names, housing, demographic, cultural will become public which violates people's privacy and contradicts with ethics.

2. Describing ethical thinking

In the book, there are 4 principles: Respect for Persons, Beneficence, Justice, and Respect for Law and Public Interest. (Salganik, 2018: p.295).

In the first comment, Kauffman (Sep. 30, 2008b) asserts that sociologists always want to get as much information of their research subjects as possible, which violates the principle of Beneficiary. According to the definition of principles of Beneficiary in the consequentialism framework, researchers "should not injure one person regardless of the benefits that might come to others" (Belmont Report 1979). Kauffman is trying to get more benefits in spite of the potential risks of leakage of the subject's information.

In the second comment, Kauffman (Sep. 30, 2008b) explains that the data he published all came from Facebook and is public. He further argues that the hackers can easily get the data he published, no need to refer to his version. This violates the principle of Respect for Persons. Although that hackers can easily get the data without the help of his data, the action of violating other's privacy is unethical or even illegal. In addition, he violates principle of Respect for Law and Public Interest in the deontology framework by show no essential respect for law and public interest (Salganik, 2018, p. 299) and Facebook's terms-of-service agreement.

In the third comment, Kauffman (Sep. 30, 2008b) argues that he has never reach out to samples for more information which corresponds to principle of Respect for Persons in the deontology framework. Because in the book, “Respect for Persons suggests that researchers should not do things to people without their consent.” (Salganik, 2018, p. 296). If he never contacted the samples, how could he get permission from the samples?

3. Ethics of Encore

a) Narayanan and Zevenbergen (2015) believe that Encore study (Burnett and Feamster, 2015) plays an important role in advancing technology in data gathering and improving ways of studying the Internet censorship of governments. However, they criticize the Encore study for some ethical issues. They suggest that the program committees should pay more attention to dealing with interactive projects (Narayanan and Zevenbergen, 2015, p.12). In the ethical issue, the authors follow the framework of "the Menlo Report" and divide the discussion to two parts: first, they analyzed who are the stakeholders of Encore; Second, they applied principle of Beneficence (Salganik, 2018, p.296). The principle of Beneficence is deeply rooted in consequentialist thinking (Salganik, 2018, p.297). That is “(1) do not harm and (2) maximize possible benefits and minimize possible harms”. (Salganik, 2018, p.297). From a consequentialism view, the authors claims that people are always at danger of digital stalking with or without Encore. However, the researchers "should not participate in and facilitate a race to the bottom"(Burnett and Feamster, 2015, p.13). In addition, the damage to users are uncertain depending on "the type of censored websites"(Burnett and Feamster, 2015, p.13)

and the responds of the censors.

b) From my perspective, the Encore study should pay more attention in research ethics and not just ethics but law should be passed to ensure to individual's information security. Based on my own experience, if chrome automatically jumped to another website, I feel being cheated and will never trust this website anymore. It not only leaves a bad user experience but also arouse anger in me.

In addition, if the websites which I am forced to connect is somehow dangerous, I will be at risk of information leakage. Nowadays, we always store passwords or credit cards in the browser. If those information was obtained by some malicious people, the result is unimaginable.

References

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.