

Assignment 8

Fulin Guo

1. a. The two examples I will describe are “Health insurance records” (Sweeney, 2002) and “Demographic, administrative, and social data about students” (Zimmer, 2010). The structure of re-identification process in these two examples is similar. We could infer people’s identity in both cases by connecting the dataset in the research with other information sources. That is, we could connect the dataset that has anonymous but sensitive information with the sources that have identity information (Salganik, 2018). In the example of “Health insurance records”, although the dataset eliminated individuals’ name, it had information about individuals’ zip, birth date, and sex. If there existed non-anonymous sources that also had this information, we could connect these non-anonymous sources with the anonymous dataset in the research to identify people. In the example of “Demographic, administrative, and social data about students”, although the Facebook data did not have participants’ name, it recorded their race and majors, so the Facebook data could be linked to specific people if we could find non-anonymous sources which also included people’s race and major information.

b. In the example of “Health insurance records”, the data GIC collected had information about individuals’ zip, birth date and sex. The data could be identified by connecting it with the non-anonymous voter registration data which also included information about individuals’ zip, birth date, and sex (Sweeney, 2002). Therefore, we could use these two datasets to deduce people’s name in the health insurance record. That is, we could infer that the two records in both datasets that had the same zip, birth date, and sex were likely to represent the same person. Then, we could link the sensitive information, like people’s diseases, in the health datasets to a specific person (Sweeney, 2002).

The re-identification process is similar in the case of “Demographic, administrative and social data about students” (Zimmer, 2010). The Facebook dataset had information about individuals’ majors, race, sex, and so on. We could first use the data of majors to identify that the source came from Harvard College because some majors listed in the data were only offered by Harvard College (Zimmer, 2010). Then, we could link the information in the code book to a specific person by connecting the Facebook data with sources that had college students’ information (non-anonymous sources), especially for some students that had unique characteristics. For instance, only one people in the codebook were Nepali (Zimmer, 2010), and if this Nepali student’s name could be found in Internet, we could confirm that the information in the Facebook data represented this student, then some sensitive information, like the student’s sexual preference, might be revealed (Zimmer, 2010).

2. I rewrite Kauffman’s comments as follows:
We have analyzed the risk of the research and considered the potential negative effects our research could cause on people. Our dataset only contains information that is on

Facebook, so the risk of the research is low since hackers are not likely to benefit from our data as they can directly find the information through Facebook as easily as using our data. (Kauffman, Sep. 30, 2008b)

We also combine consequentialism and deontology in our research. That is, we not only consider what the outcomes of our research will be, but also take people's benefits and rights into consideration. We have not asked any people questions and obtained the data that are not included in Facebook. We know it is our duty to protect people's right of privacy, so we have not published our data (Kauffman, Sep. 30, 2008b). Although it is true that there exists a possibility that someone can make extreme efforts to crack the data, we think it is difficult to do that, so the risk about our research is low (Kauffman, Sep. 30, 2008b). On the other hand, the benefits of the research are high as we can know a lot of information about individuals and this can be used to conduct social science research which will benefit the whole society. According to our risk/benefit analysis, we think it is reasonable to collect the data and conduct the research. (Kauffman, Sep. 30, 2008b).

3. a. The Encore study might not be regarded as a human-subjects research, but this research might harm people since some third party could identify who tried to visit some censored websites and punish them (Narayanan and Zevenbergen, 2015).

The beneficence principle requires researchers to consider the benefits and risks of the research. The Encore censorship study allows researchers to measure which websites are censored, and this could have a positive impact on the research in political science. The study also enables researchers to construct censorship circumvention tools (Narayanan and Zevenbergen, 2015). On the other hand, there are some potential difficulties in the research. The first is that some kinds of censorship are difficult to measure while others are easier, so the measurement outcomes might not be accurate. In addition, although the correlations are easy to detect, it is hard to confirm the causal effects of some political contexts on censorship (Narayanan and Zevenbergen, 2015).

Although the authors of the Encore censorship study say that people face the same amount of risks when they surf the Internet without Encore, (Narayanan and Zevenbergen, 2015) argues that it is hard to detect the harms of the research because trying to visit different types of censored websites might result in different consequences. Another risk is that the censors might close all Internet connections if Encore becomes very common in the world (Narayanan and Zevenbergen, 2015).

The research team should try to maximize the benefits and minimize the risks. They could reduce the risks if they got people's informed consent. However, the research team did not do so because asking for consent from people is not practical since this could largely decrease the scope of the research. Therefore, this shows that the research team cares more about the outcomes of research, not whether the methods of the research are ethical (Narayanan and Zevenbergen, 2015).

The Encore study obeys the U.S. laws, but since the research is conducted across the world, it is hard to judge whether it also obeys the laws in all countries, so the Encore study might violate some local laws (Narayanan and Zevenbergen, 2015).

b. First of all, the Encore study does not obey the principle of “respect for persons” as in the research, it does not get participants’ informal consent. Actually, people might not feel pleasant if they are aware that the study lets their computer try to visit some censored websites. Regarding the beneficence principle, the study has a lot of positive effects. It could enable researchers to measure governments censorship, study how internet censorship is implemented, and develop valid circumvention tools (Burnett and Feamster, 2015). On the other hand, it also has some risks. Some people might be punished if they tried to visit some censored websites. Whether the benefits outweigh the risks is not certain because we need more information like what positive effects the researchers have enjoyed because of this study and whether some participants have been punished due to this research. The Encore study is more likely to be based on consequentialism rather than deontology because it does not focus on participants’ autonomy as it does not obtain consent from participants. The main focus of the study is the benefits of this research (e.g. developing circumvention tool), not the means in the study (let participants’ computer visit a potentially censored website).

Reference:

Burnett, Sam and Nick Feamster, “Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests,” 2015.

Narayanan, Arvind and Bendor Zevenbergen, “No Encore for Encore? Ethical Questions for Web-based Censorship Measurement,” Technology Science, December 15 2015.
and **Vitaly Shmatikov**, “Robust De-Anonymization of Large Sparse Datasets,” 2008.

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.