Using Sensitive Data Purposefully

Authors: Kendall Darfler, Dalton George

Summary: *A computer scientist is faced with a tough decision about how to handle personally identifiable information.*

Marcus is a contract computer scientist who often works to help companies mine big data. He works with clients with diverse needs, from insurance sales to social networking. Currently, he is assisting a private research organization, the SMART Research Group, in building a database of all job-related injuries to paramedics. The SMART team is hoping to understand which types of injuries are most common, so that they can help public health officials plan injury prevention policy changes. In order to gain a comprehensive picture of all types of injuries, the team has determined that the database should include data on breaks, sprains, burns, falls, assaults and car accidents. They decided this after surveying experts in the emergency medicine field. In order to gather each of these data elements, the database will need to link data from several different resources, including ambulance companies, hospitals, and workers' compensation records. Marcus has been tasked with building an algorithm that mines and links the relevant data from each of these systems.

Based on his previous professional work, Marcus knows that this data qualifies as personally identifiable information (PII), under the United States General Services Administration Privacy Act (see Resources for Further Reading below). The U.S. Privacy Act defines PII as "information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual." Marcus knows that he is professionally obligated to protect PII. He designs an algorithm to hash all of the paramedics' identifying information, and link the data using a unique code. This way, the PII will be removed from the data set, and he will not need to access it himself.

While Marcus is working on the project, the Senior Researcher at SMART, John, talks with his friend who is a paramedic. Anecdotally, his friend mentions that he has seen quite a few needle stick injuries to his colleagues lately, and that this might be a common form of on-the-job injury for paramedics. Even though needle sticks were not determined to be significant sources of injury by the initial survey, John decides that this is an important data element to include in the job injury database. Even better, he tells Marcus, SMART has an easy way to access needle stick data for all of the paramedics: they have the city's infection control dataset for another project. They ask Marcus to help with linking this data to the injury database, and he agrees.

Marcus checks his email and sees an email from John with the subject line "Needle Stick Data." Marcus hesitates when he saw the subject line. Should he open the email?

Trusting John's judgement, Marcus clicks on the email and sees that the dataset includes PII as well as sensitive attributes, like HIV status and hepatitis infection. He feels uneasy having access to this data. He knows to look up the rules governing data use, and finds the "Minimum Necessary" Standard for Accessing Protected Health Information (see Resources for Further

Reading below). This rule states that entities should “limit unnecessary or inappropriate access to and disclosure of protected health information.” Marcus does not think that John’s attention to privacy meets the Minimum Necessary Standard, or that his attention to privacy has gone far enough. He does not feel that linking the data is the right thing to do, and it seems unnecessary given that its inclusion of the needle stick data was not precipitated by the survey. But Marcus is working for John, and John wants the data included. Pushing back against John could cost Marcus the project.

Questions:

1. There are multiple ethical issues in this case. Which ones can you identify?

2. Are there any strategies that Marcus could use to get John to reevaluate his treatment of PII?

2. How can Marcus weigh the cost of losing the project vs. acting in an unethical manner?

3. How would setting up data sharing practices at the start of the project have changed the outcome of this case?

4. Does Marcus have a responsibility to the people whose identifiers are in this city’s database, even though this data does not come from the project he has been assigned?

5. Do you handle PII or health data in your industry? What considerations should you make, in your work, for ensuring the privacy and data protection?

6. How do you address industry-specific regulations in your work?

Resources for Further Reading:

ACM Code of Ethics and Professional Conduct (see Principle 1.7 Respect the privacy of others, Principle 2.3 Know and respect existing laws pertaining to professional work)
https://www.acm.org/about-acm/acm-code-of-ethics-and-professional-conduct

Rules and Policies - Protecting PII - Privacy Act
https://www.gsa.gov/portal/content/104256

“Minimum Necessary” Standard of Accessing Protected Health Information:
http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/minimum-necessary-requirement/index.html

Kendall Darfler, MS, and Dalton George, MS, are graduates of the Drexel University Center for Science, Technology and Society. June 2017.