Introduction to Data Access and Storage

Thomas Rosenthal – DSI @ UofT

Module 06

Exit Survey: (add a link)

Beyond SQL:

Ethics

SQL in the Wild

Beyond SQL:

Ethics

SQL in the Wild

Ethics

Data Ethics issues have seen a huge increase in papers on the subject and are more and more newsworthy.

Many of these issues involve SQL databases, but not explicitly. SQL is rarely the culprit, but instead part of a system affected by *other* issues: a database full of credit cards is leaked, personal data is sold to a third party without consent, records are destroyed either intentionally or by accident.

But Data Ethics are still an important part of any well-rounded SQL course. We will consider:

- privacy and privacy policies
- security and appropriate access to data
- how database design affects our perception
- labour rights for data entry
- bias in data entry

Ethics

A Personal Story

Pakistani Computerized National ID Card

It's All People

Ethics

A Personal Story

Pakistani Computerized National ID Card

It's All People

A Personal Story

- In my first job, our most Senior Business Intelligence Analyst was involved in a motorcycle accident
- All of us were terribly concerned:
    - We had very little details of his status, other than he was in the ER and later ICU
    - We knew his 8-year old son had been involved
    - Motorcycle accidents are usually *bad*
- We were all working for the state of Colorado's largest healthcare provider, and he ended up at one of our hospitals
    - Being the organization's team of data analysts we had access to every patient's medical record

Ethics

A Personal Story

- **What are the ethical issues important to this story?**

- Privacy: we were legally and morally required to respect his privacy
  - Even if our thoughts well-intentions, we did not have his consent to these details
- Policies: his data was protected under HIPAA (PHIPA in Canada)
- Security: the organization protected his data from as many people as possible
  - We still had our job functions, requiring some access
    - The organization had logging functions in place to ensure none of us looked
  - Systems were designed to rigourously protect all patient data, including his

---

- He lived, with a long road to recovery
- His son had very few injuries and recovered quickly
- None of us looked at any of his records

Pakistani Computerized National ID Card

- Read: Qadri, R. (2021, November 11). *When Databases Get to Define Family.* Wired.
  https://www.wired.com/story/pakistan-digital-database-family-design/

Ethics

Pakistani Computerized National ID Card

- **What values systems are embeddedd in databases and data systems you encounter in your day-to-day life?**

- Fairness: Our expectations of what is "normal" can inform software and database design
  - Many studies show that lack of diversity in tech creates a narrow viewpoint
    - Design choices can reinforce harmful stereotypes, be exclusionary, further marginalize people
    - Shifting norms...?
- Inequality: Data and data systems are representation of history, but influence the present and the future
  - This effect is compounded when we add automation to these data systems
    - These systems often claim to address bias and inequity, but are better at disguising it
    - Allow us to escape the difficulties of making decision about actual people
  - What gets counted? Who gets counted?

It's All People

- Read: Boykis, V. (2019, October 16). *Neural nets are just people all the way down.* Normcore Tech.
  https://vicki.substack.com/p/neural-nets-are-just-people-all-the

Ethics

It's All People

- **What are the ethical issues important to this story?**

- Labour: like Machine Learning, SQL databases are built on a foundation human labour
    - Much of this labour is invisible
        - I was aware that much of the health data I used was produced by "coders", but generally only thought about these people when there were mistakes
- Bias: Humans make choices, labels, determinations before a neural network ever exists
    - Humans can create bias and are subject to certain pre-conceptions (i.e. purposes for labelling) that will create different outcomes
        - Biases have to be addressed/moderated...more human labour!

Beyond SQL:

Ethics

SQL in the Wild

SQL in the Wild

For me personally, I love using Firefox's built in SQLite DB "places"

1. Navigate to `about:profiles` in Firefox
2. Open your root director profile using the Show in Finder/Explorer window (e.g. /Users/thomas/Library/Application Support/Firefox/Profiles...)
3. Make a copy of `places.sqlite` (I usually just save to desktop)
4. Open the copy in DB Browser for SQLite

- Which website do you visit most often (`moz_history`)?
- What do you type most often into the search bar (`moz_inputhistory`)

Lingering Questions?