CASE STUDIES

Ensuring Data Privacy in Large-Scale Data Sharing

Evîn Cheikosman

To illustrate the feasibility and effectiveness of implementing a city controlled data privacy system, we can draw insights from 2 significant initiatives: the US Census Bureau's use of differential privacy in the 2020 Census and Los Angeles' GeoHub platform. These examples demonstrate successful strategies for protecting sensitive data while ensuring necessary access, providing valuable lessons for the proposed API wrapper system in San Francisco.

1. US CENSUS BUREAU: DIFFERENTIAL PRIVACY

Context

The Census Bureau faced the challenge of releasing detailed demographic data while ensuring that individuals' identities were protected. To address this, the bureau implemented differential privacy for the 2020 Census. Differential privacy is a statistical technique that introduces controlled noise into the data, making it difficult to re-identify individuals from published datasets. This approach was designed to balance the need for data accuracy with robust privacy protection.

Relevance to San Francisco

The Census Bureau's adoption of differential privacy is directly relevant to San Francisco's efforts to protect sensitive healthcare data, particularly in the context of reverse warrants. Similar to the challenges faced by the Census Bureau, San Francisco must ensure that data shared with law enforcement under reverse warrant requests do not inadvertently expose sensitive personal information.

> *"The challenge of balancing data utility with privacy protection is one that no organization releasing information drawn from confidential or sensitive databases can afford to overlook. This is especially true when dealing with very sensitive information like healthcare data. Our work at the US Census Bureau has shown that investing in advanced privacy systems is not just about protecting data; it's about building trust for both those who respond to our surveys and our users who rely on accurate data."*
>
> **Ron Jarmin, Deputy Director and Chief Operating Officer, US Census Bureau**

Application

The principles of differential privacy could inform the design of San Francisco's API wrapper system. By introducing techniques that obscure specific data points while preserving the overall utility of the data, the city can protect individuals seeking reproductive and gender affirming healthcare services. This approach ensures that law enforcement can still access necessary data for legitimate investigations without compromising the privacy of innocent individuals.

2. LOS ANGELES'S GEOHUB: SECURE GEOSPATIAL DATA SHARING

Context

Los Angeles's GeoHub platform serves as a centralized repository for the city's geospatial data, providing access to a wide range of datasets for public use, business applications, and government operations. Given the sensitive nature of location data, the GeoHub has implemented stringent measures to protect individual privacy while still making valuable data available for public use. This includes anonymizing location data and carefully controlling access to more sensitive datasets.

Relevance to San Francisco

The GeoHub's approach to managing and sharing sensitive geospatial data offers valuable lessons for San Francisco as it seeks to develop its own API wrapper system. Like Los Angeles, San Francisco must balance the need to share data with law enforcement against ensuring that sensitive information, particularly related to healthcare, is not exposed.

APPLICATION

San Francisco can model its API wrapper system on the privacy protocols employed by GeoHub. This might involve anonymizing or aggregating data before they are shared with law enforcement and implementing strict access controls to ensure that only authorized individuals can view the data. By leveraging the successful strategies used by GeoHub, San Francisco can create a robust system that protects sensitive healthcare data while still complying with legal obligations.

CONCLUSION

These case studies from the US Census Bureau and Los Angeles' GeoHub demonstrate that it is possible to manage and share large datasets while protecting individual privacy. By drawing on these examples, San Francisco can implement a city controlled data privacy system that not only meets legal and operational needs but also safeguards the privacy of its residents. This approach would enable San Francisco to uphold the principles of the recently passed San Francisco Reproductive Freedom Act, ensuring that sensitive healthcare data are protected in the digital age.