

Thesis Project Portfolio

Emergency Management and Underserved Communities: Using Big Data to Improve Emergency Management Preparedness, Response and Resilience

(Technical Report)

Privacy in the Digital Age: Ethical data collection and storage systems for use in the public domain

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Zachery Key

Spring, 2022

Department of Systems Engineering and Environment

Table of Contents

Sociotechnical Synthesis

Emergency Management and Underserved Communities: Using Big Data to Improve Emergency Management Preparedness, Response and Resilience

Privacy in the Digital Age: Ethical data collection and storage systems for use in the public domain

Prospectus

Sociotechnical Synthesis

Current Hurricane evacuation techniques rely on a complex sequence of events that begin with an extreme weather forecast issued to the state governments by federal agencies (e.g., National Weather Service, National Hurricane Center) and culminate in state or local governments issuing an evacuation order. Because modeling the uncertainty of a hurricane strike is an inverse function of time, governing bodies must decide early on whether to risk the lofty expenditures of a false alarm or wait and risk endangering lives. Too many false alarms often result in a decreased adherence to evacuation policy and result in less people evacuating even when a mandatory evacuation order has been issued.

In order to produce a better understanding of constituent's evacuation patterns, my capstone team investigated a number of factors that are involved in an individual's decision to evacuate. Building upon previous studies, our team utilized smart phone mobility data collected from Cuebiq to model evacuation patterns in a 3-month period centered around hurricane evacuation orders issued in VA, NC and SC during hurricane Florence (2018). We then joined this data with census block level demographic data, evacuation zone polygons and evacuation order data to characterize trends in evacuation behavior.

Despite the rapid growth of data collection and storage system over the past 2 decades, very little has changed in the world of data anonymization and obfuscation. This presents a very real threat to our personal security if the databases that hold our information are compromised. Moreover, if you can reverse engineer these data points to identify their source signal transmitter, the same points could be traced to reveal visitation patterns to locations such as political rallies, nightclubs, marijuana dispensaries, abortion clinics, LGBTQ centers or rehabilitation centers. Without ever meeting you in real life, a stranger can uncover highly sensitive personal

information such as your political affiliation, sexual orientation, substance consumption, and/or nightly routines.

In order to ethically utilize the mobility data in statistical analysis, participants privacy must be protected. In order to determine if the Cuebiq data provides a reasonable level of privacy to the people whose data is being collected, I compare their data structure to that of the far more controversial mobility database structure found in the Los Angeles Department of Transportation's Mobility Specification (LADOT's MDS). Drawing from Star's theory of the infrastructure of technology, I will investigate the embedded values of privacy and utility found in the two systems as a function of their respective database designs. By comparing the spatiotemporal resolution alongside the reidentification potential of individual records contained in the two databases, I identify the valuation of privacy by each organization and consider the tradeoff between privacy and utility.