CODIS AND PROBABILISTIC GENOTYPING: NAVIGATING SUCCESS AND SECURITY

STANDARDS OF USE OF DNA DATABASES: PRIVACY AND PUBLIC INFORMATION

An Undergraduate Thesis Portfolio
Presented to the Faculty of the
School of Engineering and Applied Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Lily Roark

May 9, 2022

SOCIOTECHNICAL SYNTHESIS

In criminal forensics, the results from deoxyribonucleic acid (DNA) analysis and Combined DNA Index System (CODIS) database entry provide new leads for law enforcement where a case might otherwise have gone cold, and are crucial pieces of evidence in criminal court. It is imperative that the personal biological data within CODIS is not misinterpreted or abused. The technical report outlines the publicly available security measures used in CODIS, and where the Federal Bureau of Investigation (FBI) has kept its measures secret, recommends secure software design paradigms. Intrinsically coupled to this technical topic in both subject and motive, the science, technology and society (STS) report scrutinizes how to expand standards of use of DNA databases to protect public privacy.

The technical report synthesizes the computer science subfields of databases, algorithms, and cybersecurity as they apply to CODIS. There are massive backlogs of DNA samples awaiting CODIS entry, and developments in probabilistic genotyping and automated DNA entry tools exert pressure to securely integrate with the old binary-matched system. Vulnerabilities introduced by integration require methodological barriers preventing malicious parties from accessing CODIS data; these barriers need to be scalable and efficient, as the database is ever-growing and the methods for analyzing its records become more computationally complex. Research taken from the FBI, legislative proceedings, leading forensic scientists, and security-focused academics comprise the CODIS security overview and recommendations.

Given this research, the technical report finds that CODIS is already extremely secure at multiple levels of database security. The records contained in CODIS do not reveal anything about the individual's phenotype (physical qualities), and to retrieve the name that the record originated from, a match must be cross-referenced analyzing lab. The servers containing the data

are physically secured within FBI headquarters, there are firewalls to prevent network intervention, and an unspecified encryption of the data within the CODIS application. The report recommends specifically AES encryption, and that the database be continually updated against evolving adversaries.

The STS report questions how additional standards of forensic DNA databasing should be set in order to maintain public transparency and avoid malpractice. Sources for addressing this problem include FBI documents, conclusions of forensic conferences, and analogous STS case studies. The analysis-guiding framework is Actor Network Theory (ANT) as defined by Callon and Law, which maps out forces influencing the issue and their relationships to each other. Ultimately, standards must be extended beyond the limited definition of technology, the laboratory which analyzes DNA samples, to include other actors in the ANT model.

In regards to CODIS, the ANT actors can be clustered into categories of the judiciary, the development of the technology, the laboratory, and the sample data collection. Analogous case studies show that each group is vulnerable in different ways: the judiciary misinterprets the technology, the developers lack oversight, and the labs cannot process records fast enough. In order to overcome the lag of regulation and standardization of DNA databasing, these groups need to be in constant dialogue with each other. Through constant communication, these groups can remediate the vulnerabilities within a given sector.

Together, the technical and STS reports examine the current public measures to ensure CODIS as a system is private and operating efficiently, then ways to maintain privacy and efficiency in light of developing DNA technology. The expanding maintenance should include continuous discussion across fields of expertise, and critical evaluation of the database infrastructure.

TABLE OF CONTENTS