

Introduction to Data Science
March 2, 2022
Marissa Beaty
Assignment 1

The dataset I am proposing as a contribution to Mimi Onuoha's "Missing Datasets" collection is a list of individuals whose data has been pulled or reviewed by the National Security Agency (NSA) in the United States and why. This dataset differs from Onuoha's in that it exists, but I am proposing its addition because it exists under one sided control and access. To understand why this dataset is not publicly available and how that implicates and perpetuates imbalanced power structures, we must first understand who the National Security Agency is and what they do.

The National Security Agency was founded in 1952 by U.S. President Harry S. Truman from a World War II operating unit tasked with deciphering coded communications. Today, its mission involves the “global monitoring, collection, and processing of information and data for foreign and domestic intelligence and counterintelligence purposes” (NSA, 2023). In 2013, the NSA was studied and scrutinized for its unethical data collection processes, which were brought to light by former employee Edward Snowden. Snowden revealed the NSA “intercepts and stores the communications of over a billion people worldwide, including United States Citizens” (The Guardian, 2013). In 2020, the practices brought to light by Snowden were ruled unlawful and a violation of U.S. citizens right to privacy (U.S. Court of Appeals, 2020).

From this brief introduction, the NSA has already established itself as 1) a force of power previously unchecked by other government agencies and 2) using illegal practices to obtain data from unknowing U.S. and non-U.S. citizens. The 2020 court case ruled on the legal implications of the NSA's data collection processes but does not explain who and what this dataset included, how long the data was stored, nor why the data was collected in the first place. This has already established an unjust power structure between the NSA and the global civilians who had their data unrightfully taken. Exacerbating this power structure is the lack of information provided by the NSA regarding those its unlawful practices affected. Without knowing who is in this dataset, global citizens must live in constant fear that they are being “watched.”

We must ask ourselves, if the dataset was created using illegal practices, why does the dataset exist at all? And why is it not made public as I am suggested it be? This goes back to the foundation of the NSA as a domestic intelligence and counterintelligence services. As they

argued in the 2020 court case, the data they collected was done so under the protection of The Foreign Intelligence Surveillance Act, Subchapter 1, which authorizes the “electronic surveillance” of foreign agents and U.S. targets when those agents and targets have demonstrated “probably cause” to be involved in acts of domestic and international terrorism (U.S. Department of Justice). As the Case ruled, this Act does not hold up for the nearly one billion individuals whose data was collected and stored by the NSA. To put it simply, the NSA is a powerful organization imposing an unequal power structure on global citizens.

My proposed dataset tries to invert that, or at minimum, equalize the power balance between the NSA and the people its searches have affected. Releasing the names of those inquired about and stored by the NSA creates transparency and allows those inflicted to reclaim their data and act upon having their information illegally acquired. This is a drastic improvement from the current system which requires everyone to assume their data can be accessed at any given time (or has already been accessed), and the fear that instills. It would be naïve to forget how this might impact those who pose real threats to national security and how this transparency might allow them to create alternate tactics to evade detection on their unlawful endeavors. Despite this, the benefits of releasing this dataset support my conclusion that it should exist publicly. Though there is the concern of the few who remain threats to national security, the dataset, as it stands, exists illegally, and whatever it contains had no right being collected in the first place. Furthermore, by maintaining the secrecy of this dataset, anyone who has ever accessed the internet or used a mobile device must live in fear that their data has been collected without their consent. Should the dataset be released, the constant fear of the many, can be reduced to the few: those whose data was collected and stored. In doing so, it also allows those few an opportunity to reclaim their data. Given these reasons, this dataset is a natural addition to Mimi Onuoha's work.

Works Cited

Bureau of Justice Assistance (n.d.). *The Foreign Intelligence Surveillance Act of 1978 (FISA)*. [online] Bureau of Justice Assistance. Available at: <https://bja.ojp.gov/program/it/privacy-civil-liberties/authorities/statutes/1286>.

Greenwald, G., MacAskill, E. and Poitras, L. (2013). *Edward Snowden: the whistleblower behind the NSA surveillance revelations*. [online] The Guardian. Available at:

<https://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance>.

Wikipedia Contributors (2019). *National Security Agency*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/National_Security_Agency.