

LEGAL & ETHICAL ANALYSIS OF RE-IDENTIFICATION IN OPEN DATASETS

City of Boston's Annual Employee
Earnings Report

Ferdous Alam, Sissie Cui, Kelly Nguyen



THE DATASET

EMPLOYEE EARNINGS REPORT

Each year, the City of Boston publishes payroll data for employees. This dataset contains employee names, job details, and earnings information including base salary, overtime, and total compensation for employees of the City.

See the "Payroll Definition Categories" document below for an explanation of what types of earnings are included in each category.

[↗ Embed](#)

Add Filter

Show entries

Showing 1 to 10 of 22,552 entries

[Hide/Unhide Columns](#) [Download](#)

Search:

NAME	DEPARTMENT_NAME	TITLE	REGULAR	OTHER	OVERTIME	TOTAL_GROSS	POSTAL
Beckers,Richard	Boston Police Department	Police Officer	None	1,264,843.63	None	1,264,843.63	02119
McGowan,Jacqueline M.	Boston Police Department	Police Officer	None	1,252,990.81	None	1,252,990.81	02129
Harris,Shawn N	Boston Police Department	Police Offc Comm Serv Offc 358	69,772.10	212,739.48	82,300.87	433,073.75	02130
Washington,Walter	Boston Police Department	Police Officer	100,963.38	211,900.28	67,849.66	399,825.87	02368

Employee Earnings Report 2021

One of the open datasets hosted on Analyze Boston, Boston's Open Data Hub

REFRESHED ANNUALLY

Each year, the City of Boston **publishes payroll data** including employee names, job details, and earnings information including base salary, overtime, and total compensation for employees of the City






12 KEY COLUMNS

NAME, DEPARTMENT_NAME, TITLE, REGULAR, RETRO, OTHER, OVERTIME, INJURED, DETAIL, QUINN_EDUCATION_INCENTIVE, **TOTAL_GROSS**, **POSTAL**

PUBLISHED BY ANALYZE BOSTON

This new open data platform is a part of the city of Boston's **Open Data to Open Knowledge project**

OUR RE-IDENTIFICATION RESULTS

Name	Title	Address	Total Compensation	Home Value	Phone
	Police Offc Comm Serv	<u>27 Sanders Dr, Saugus, MA 01906</u>	\$218,896.36	\$822,827	781-558-5150
	Captain/Academy Instructor	<u>74 Ridge Hill Dr, Hanover, MA 02339</u>	\$291,315.37	\$536,000	781-982-9466
	Police Captain/DDC	<u>60 Clifftondale St, Roslindale, MA 02131</u>	\$336,410.11	\$975,600	617-962-4865
	Head Trainer	<u>28 Wilcock St, Dorchester Center, MA 02124</u>	\$110,454.96	\$429,900	617-436-0487
	Police Lieutenant	<u>10 Clover Cir Melrose, MA 02176</u>	\$305,741.24	\$983,500	617 593-0814

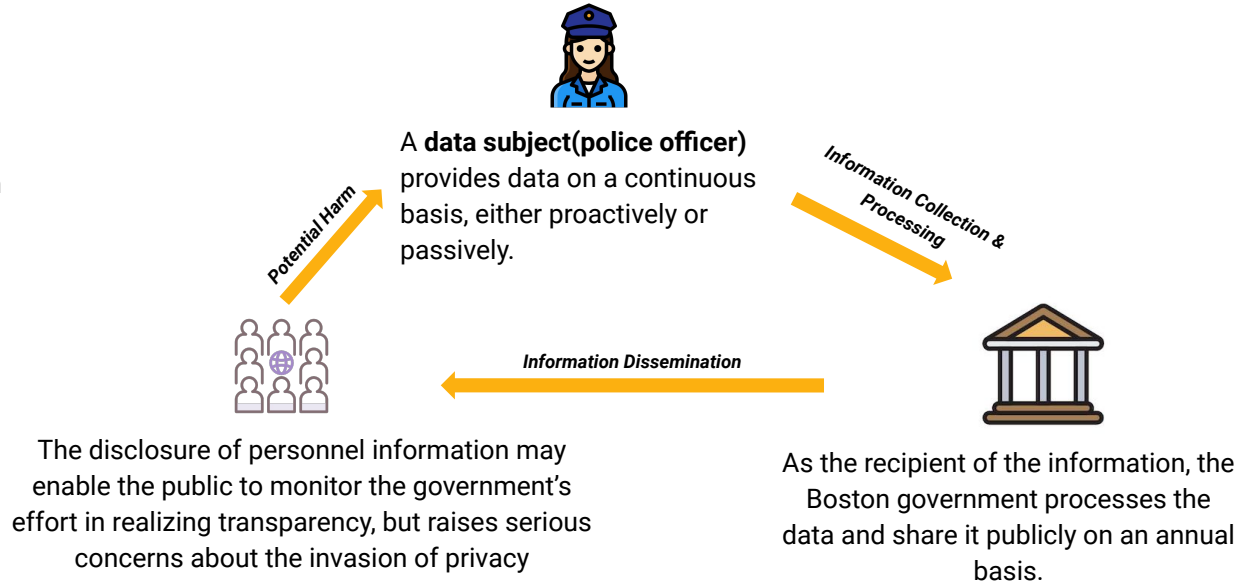
PRIVACY ANALYSIS

Taxonomy and Contextual Integrity

Context

- A major purpose of the Public Records Law is to enable taxpayers to monitor government activities, expenditures and employees.
- Disclosing detailed personal information seems to be understandable despite concerns about invasion of privacy

How is the data collected, processed and disseminated?



PRIVACY ANALYSIS

Mulligan et al.'s analytic



Data openness and privacy are fundamentally **contrast concepts**



Privacy is not **justified** because personal identifiable information is disclosed to the public and more personal information such as home address could be further identified



Potential **harms** include physical safety and financial fraud and scams (i.e. credit card related fraud)



Due to the Public Record Law, as an employer, the Boston government has no **mechanism** to protect its employee, the police officer's privacy



A full disclosure of detailed personal information seems to be a **social practice** in the context of government transparency



The employee records of the past 10 years are still available today. **Data minimization** and **data retention** practices need to be reviewed to avoid unnecessary disclosure

ETHICAL & LEGAL IMPLICATIONS

First Name & Last Name



IT'S A KEY IDENTIFIER

First and Last Name is the key identifier for multiple online databases, social media, government look-ups (criminal history, sexual offenders)



ARE WE DOOMED TO REPEAT HISTORY?

An ethical dilemma exists: Sensitive data may adversely affect one's reputation. Just because a crime has been committed, it is not indicative of all future actions

Zip Code



PAIR NAME + ZIP TO GET EVEN MORE INFORMATION

Pairing the two will get almost exact location, eventually homeowner, tax information, etc.



POLICE OFFICERS ARE ESPECIALLY VULNERABLE

Criminals may behave vindictively, actively trying to re-identify the open dataset in order to invade privacy or cause bodily harm

CURRENT LAW AND OUR RECOMMENDATIONS

MASSACHUSETTS PUBLIC RECORDS LAW



- Every government record in Massachusetts is **presumed to be public** unless there is an exemption
- **Anyone may request** records directly from a Records Access Officer (RAO)



OPEN DATA DOES NOT ABIDE BY THE SAME LAW

- There is **only vague guidance** for the open datasets in Boston's *Open and Protected Data Policy*
- Open data is especially sensitive because anyone can access at any time and **there is no request record**



STEPS TO IMPROVE

- Open Data should ideally be governed by **stricter** Code of Conduct
- Take smaller steps to improving privacy: (1) learning the potential positions of harm (2) create a request form as a way to create a moderate barrier to entry to deter malicious intents
- Data minimization and perturbing to the data is recommended

THANK YOU!

ANY QUESTIONS 