The City of New York
Department of Investigation

JOCELYN E. STRAUBER
COMMISSIONER

180 MAIDEN LANE
NEW YORK, NY 10038
212-825-5900

Release #31-2024
nyc.gov/doi

FOR IMMEDIATE RELEASE
THURSDAY, JULY 11, 2024

CONTACT: DIANE STRUZZI
(212) 825-5931

DOI ISSUES ANNUAL ANTI-CORRUPTION REPORT FOR 2023 FOCUSED ON THE INTEGRITY OF CITY DATA AND ITS UTILIZATION IN AGENCY ANTI-CORRUPTION EFFORTS

Today, Jocelyn E. Strauber, Commissioner of the New York City Department of Investigation ("DOI"), issued the 2023 Annual Anti-Corruption Report on the topic of data integrity — how agencies ensure the accuracy and consistency of their data — and agencies' use of their data to address fraud, waste and corruption risks. Data integrity is defined as the quality, accuracy, consistency and security of data, the verification of its accuracy and consistency maintained over time and across formats, and the enforcement of rules and standards that prevent unauthorized data alteration. Data integrity is crucial because it ensures the trustworthiness and reliability of data, enabling informed decision-making and efficient operations, and strengthens data security by controlling access and preventing misuse. To the extent agencies use their data in their anti-corruption efforts, data integrity is one component of the effectiveness of those efforts.

DOI's Annual Anti-Corruption Report is mandated by Executive Order 105 ("EO 105"), which consolidated the Inspector General function within DOI and established the DOI Commissioner as the City's independent Inspector General, but gave agency heads primary responsibility for maintaining corruption-free agencies, and called upon DOI to assist in their efforts by preparing this annual Report, which summarizes agency-identified corruption vulnerabilities and agencies' remedial strategies. Since 2020, these annual reports have focused on City agencies' responses to a corruption-related issue. Unlike other DOI Reports, these annual reports rely primarily on information and analysis supplied to DOI by City agencies, rather than DOI's own investigative work.

For this Report, DOI analyzed questionnaire responses from 48 agencies. The Report covers the period October 1, 2022 through September 30, 2023. A copy of the Report is attached to this release and can be found on DOI's Reports page or by clicking here.

DOI Commissioner Jocelyn E. Strauber said, "Maintaining accurate and consistent data and using that data in anti-corruption efforts, are significant responsibilities of every New York City agency. Data, and therefore data integrity, impacts the City's decision-making, record-keeping, and how agencies serve the public. This Report found that the majority of agencies use data to combat corruption and have practices or written policies designed to protect data integrity, but also found areas where agencies could improve. To that end, DOI recommended that City agencies review their data integrity practices and consider improvements such as memorializing practices in writing, controlling database access, and testing recovery procedures. I thank all the City agencies that responded to our questionnaire for their participation in the creation of this Report and for their commitment to maintaining data integrity."

more

This Report relies on agencies' own assessments to provide a broad overview of the approach City agencies are taking to address risks of corruption, misconduct, or other criminal activity. In order to promote candid responses by the participating agencies, the individual responses have been aggregated or anonymized, as appropriate. For this 2023 Report, DOI developed a questionnaire that probed agencies' approaches to data integrity by posing questions in the following categories:

- **Identification** – agencies were asked to identify and describe all databases that they control, directly or via contract, for which that agency is the primary user; what type of data each database contains; who owns the database; whether the data is captured manually or electronically; and whether the data is stored on-premises or in the cloud;

- **Risk Analysis** – agencies were asked to identify the five databases they deem most critical, to describe the data integrity measures in place for those databases, to identify any data integrity risks not addressed by those measures, and whether the agency has written data integrity policies or procedures; and

- **Proactivity** – agencies were asked whether they use any of their databases for purposes of identifying, preventing, or mitigating risks of corruption, fraud, waste, or abuse; whether they use artificial intelligence for anti-corruption efforts; and whether agencies have staff or units specifically responsible for analyzing, monitoring, or auditing data contained in those databases.

Data integrity is conceptually distinct from cybersecurity, which was the focus of DOI's 2021 Anti-Corruption Report. Cybersecurity focuses on protecting systems and networks from digital attacks from both external and internal threats. Data integrity refers to preserving the validity and accuracy of data largely from internal threats that can involve internal bad actors or accidental human error. Inaccurate data can lead to poor decision-making and thus government waste; unmonitored data may present an opportunity for internal bad actors to commit acts of corruption and fraud. Proper data integrity policies can help avoid both.

City agencies' questionnaire responses indicate that most agencies have taken steps to protect the integrity of their data, but the responses also exposed areas in which agencies could improve. The Report found that the majority of City agencies utilize some combination of data integrity best practices, including user-based limitations on access, audit trails, and periodic internal audits. However, agencies self-reported key risks to data integrity. Those risks included lack of role-based permissions and access, increased employee turnover and retention issues, as well as the advanced age of certain database platforms and maintenance of those platforms by third-party, outside vendors.

Thirty agencies indicated that they had written policies or procedures in place governing data integrity, though five of those agencies appear to rely solely on Citywide policies, rather than agency-specific policies on data integrity. A few of those thirty agencies submitted audit policies that did not explicitly mention or address data. Eighteen agencies reported they had no written policies on data integrity, though each of those agencies reported having practices in place to address data integrity—presumably, then, not memorialized in writing.

Thirty-three agencies responded that they do use data to identify, prevent, reduce, or eliminate instances or risks of corruption, fraud, waste, or abuse. Agencies reported methods including cross-referencing datasets against each other to flag issues. Six of the 48 agencies responded that they use artificial intelligence to analyze, monitor, or audit data to prevent corruption, fraud, waste, or abuse. These agencies reported a variety of helpful AI uses, such as algorithm-based evaluations of transaction characteristics to assign a fraud risk score for credit card authorization requests.

Forty agencies had staff responsible for data integrity efforts; eight agencies reported having no staff or units responsible for analyzing, monitoring, and/or auditing data contained in any of the identified databases. Of the 40 agencies with staff, headcounts varied but appeared largely proportional to the size of the particular agency. The experience of those staff members also varied and included a mix of advanced degrees, civil service exams or certifications, and general expertise in the data system itself or experience in City government.

Based on these findings, the Report recommends that City agencies assess their current data integrity policies and practices to evaluate whether they adequately promote data integrity and sufficiently utilize data to address risks of fraud, corruption, and abuse. As part of that assessment, DOI issues the following five recommendations for agencies to consider in light of their specific needs:

- Ensure that the agency has a written data policy that includes provisions regarding data governance, such as access control and disaster recovery procedures. Such data policy should be periodically reviewed and updated as necessary.
- Appoint a data officer responsible for setting the data policy, determining access, and reviewing compliance.
- Where possible, phase out the manual entry of data, moving to electronic input only. With respect to data deletion, limit deletion authority to a small universe of appropriate supervisory staff.
- Control database access based on roles or groups to which individuals are assigned so that access is consistent across similarly situated staff.
- Test and simulate disaster recovery procedures periodically to ensure that they will work as intended when actually needed.

This Report was prepared by Deputy Commissioner of Legal Affairs and General Counsel Andrew Brunsden, New York City Legal Fellow Sophia Khorshad, Director of Intergovernmental Affairs and Special Counsel Rebecca Chasan, and Director of Data Analytics Shyam Prasad, and was supervised by Deputy Commissioner of Strategic Initiatives Christopher Ryan and Deputy Commissioner/Chief of Investigations Dominick Zarrella.

New York City
Department of Investigation

2023 Annual
Anti-Corruption Report

Jocelyn Strauber
Commissioner

Table of Contents

I. Introduction

The Commissioner of the Department of Investigation ("DOI") serves as the City's independent inspector general, supervising a staff of Inspectors General, investigators, attorneys, forensic auditors, computer forensic specialists, and administrative personnel. Mayoral Executive Order No. 105 (EO 105)[1] established this structure in 1986, consolidating the Inspector General functions of various City agencies within DOI.

EO 105 also made clear that agency heads "remain principally responsible for maintaining corruption-free agencies through this formal collaborative arrangement by developing procedures and systems to protect against corrupt and other criminal activity affecting their agency, by hiring employees of integrity and competence, by careful managerial oversight and high-quality supervision of agency employees, and by adequate review and monitoring of fiscal commitments and processes within their respective agency." This mandate also is reflected in New York City Charter § 389(a), which assigns to agency heads the responsibility for maintaining "an internal control environment and system which is intended to maximize the effectiveness and integrity of agency operations and to reduce the vulnerability of the agency to fraud, waste, abuse, error, conflict of interest, and corruption."

EO 105 calls upon DOI to assist agencies in their efforts by preparing this Annual Anti-Corruption Report. The Annual Anti-Corruption Report is intended to describe the corruption hazards identified at City agencies and the strategies identified by agencies, in consultation with the various Inspectors General at DOI, to address those hazards. Although EO 105's mandate applies only to Mayoral agencies, DOI invited non-mayoral agencies to participate in this report and seven non-mayoral agencies contributed, as well as 41 mayoral agencies.

DOI's anti-corruption report, unlike DOI's other reports, is not the product of DOI's investigative work, but rather provides a broad overview of approaches taken by City agencies to address risks of corruption, misconduct, or other criminal activity. DOI generally obtains this information by submitting questions to other City agencies and analyzing their responses. The public release of this report is intended to promote information-sharing and transparency around corruption hazards and the measures that are being used by City agencies to combat corruption. In order to promote candid responses by the agencies participating in this and future questionnaires, the individual responses have either been aggregated or anonymized, as appropriate.

[1] Available at: https://www.nyc.gov/assets/records/pdf/executive_orders/1986EO105.PDF.

II. Executive Summary

The focus of the 2023 Anti-Corruption Report (the "Report"), which covers the period of October 1, 2022 to September 30, 2023, is the integrity of City data and its utilization in agency anti-corruption efforts. The goal is to understand how agencies ensure the accuracy and consistency of their data and to determine whether agencies utilize data to address risks of fraud, waste, and abuse. Data integrity refers to the quality, accuracy, consistency, and security of data, the verification of its accuracy and consistency maintained over time and across formats, and the enforcement of rules and standards that prevent unauthorized data alteration.[2],[3] Data integrity is crucial because it ensures the trustworthiness and reliability of data, enabling informed decision-making and efficient operations, and strengthens data security by controlling access and preventing misuse.[4]

New York City agencies are inundated with large quantities of data that can be exploited or abused if proper data integrity measures are not in place. To evaluate the state of City agencies' data integrity, DOI developed a questionnaire (the "Questionnaire" or "DOI's 2023 Anti-Corruption Questionnaire") that sought responses from agencies about three categories of relevant information for an assessment of data integrity: identification, risk analysis, and proactivity.

Data integrity can include both physical data integrity and logical data integrity. Physical data integrity refers to the physical precautions and risks associated with data, such as preventing equipment damage and creating safeguards against physical threats such as power outages and storage erosion.[5] Logical data integrity refers to the ability to keep data substantively consistent and accurate over long periods of time. Logical data integrity includes preventing duplicates or null values, storing data uniformly and consistently across a given database, and limiting data creation and modification rights to enumerated users.[6] All of these data management elements help ensure logical, or substantive, data integrity is protected.

Data integrity is conceptually distinct from cybersecurity, which was the focus of DOI's 2021 Anti-Corruption Report. Cybersecurity focuses on protecting systems

[2] *What is Data Integrity?*, IBM (April 5, 2024), https://www.ibm.com/topics/data-integrity. *See also* Catherine Cote, *What Is Data Integrity and Why Does It Matter?,* Harvard Business School Online (Feb. 4, 2021), https://online.hbs.edu/blog/post/what-is-data-integrity.

[3] While preparing this Report, DOI did not locate citywide policies or procedures dedicated solely to data integrity. DOI turned to reputable industry sources to help define and establish common industry terms.

[4]. *What is Data Integrity?*, IBM (April 5, 2024), https://www.ibm.com/topics/data-integrity.

[5] *See id.*

[6] *See id.*

and networks from digital attacks from both external and internal threats.[7] Conversely, data integrity refers to preserving the validity and accuracy of data largely from internal threats that can involve internal bad actors or accidental human error.[8] Inaccurate data can lead to poor decision-making and thus government waste; unmonitored data may present an opportunity for internal bad actors to commit acts of corruption and fraud. Proper data integrity policies can help avoid both. Governments and institutions have acknowledged the significance of data, and therefore of data integrity, in efforts to combat fraud and corruption.[9] The Questionnaire probed agencies' approaches to data integrity by posing questions in the three above-referenced categories: identification, risk analysis, and proactivity.

Identification. The Questionnaire asked agencies to identify and describe all databases that they control, directly or via contract, for which that agency is the primary user. Agencies were asked to identify what type of data each database contains, who owns the database, whether the data is captured manually or electronically, and whether the data is stored on-premises or in the cloud.

Risk Analysis. The Questionnaire asked agencies to identify the five databases they deem most critical and describe the data integrity measures in place for those databases. The Questionnaire then asked agencies to identify any data integrity risks not addressed by those measures and probed whether agencies had written data integrity policies or procedures.

Proactivity. The Questionnaire also asked agencies whether they use any of their databases for purposes of identifying, preventing, or mitigating risks of corruption, fraud, waste, or abuse. This section of the Questionnaire inquired about agency use of artificial intelligence for anti-corruption efforts and whether agencies

[7] Catherine Cote, *What Is Data Integrity and Why Does It Matter?,* Harvard Business School Online (Feb. 4, 2021), https://online.hbs.edu/blog/post/what-is-data-integrity; *see also Data Protection 101*, Data Insider (May 8, 2023), https://www.digitalguardian.com/blog/what-data-integrity-data-protection-101.

[8] Catherine Cote, *What Is Data Integrity and Why Does It Matter?,* Harvard Business School Online (Feb. 4, 2021), https://online.hbs.edu/blog/post/what-is-data-integrity.

[9] For instance, the World Bank has built an online application to detect patterns among companies bidding on World Bank-financed projects.[9] Ethiopis Tafara, *How the World Bank Is Using Data To Better Detect Fraud and Corruption*, World Bank Blogs (Jan. 21, 2020), https://blogs.worldbank.org/governance/how-world-bank-using-data-better-detect-fraud-and-corruption. In Brazil, an algorithm was created capable of using open data related to the Brazilian Congress' reimbursement quotas and identifying legitimate or suspicious expenses. *Can Data and AI Be Used as a Weapon to Fight Corruption?*, Data-Pop Alliance (June 9, 2021), https://datapopalliance.org/lwl-28-data-and-anti-corruption/. In India, an initiative called "I Paid A Bribe" tackles corruption by maintaining a website where anyone can report the nature, number, pattern, types, location, frequency, and value of bribes they have paid to government officials. *Available at*: https://www.ipaidabribe.com/#gsc.tab=0.

have staff or units specifically responsible for analyzing, monitoring, or auditing data contained in those databases.

City agencies' Questionnaire responses indicate that most agencies have taken steps to protect the integrity of their data. Responses also exposed areas in which agencies could improve. The Identification section tasked agencies to consider important aspects of these databases which, in turn, allowed agencies to consider the impact of database attributes on risk analysis. Responses to the Risk Analysis portion of the Questionnaire revealed some agencies were doing better than others at addressing data risks. For instance, some agencies who noted platform age or vendor as a risk also noted efforts already underway to upgrade systems and replace vendors. And finally, some agencies identified proactivity strategies in a variety of different forms, to be explored in this Report. For instance, one agency established and maintains a Data Governance Advisory Board.

While each City agency has a different mission, all agencies maintain important data. This Report offers a useful description of agency efforts to protect the integrity of this data and use it to address corruption risks.

III. Questionnaire Results

A. Respondents

There were 48 respondents to DOI's 2023 Anti-Corruption Questionnaire. Seven of the 48 respondents were non-mayoral agencies. Table 1 below lists the agencies and offices that submitted a response.

Table 1. Respondents to DOI's 2023 Anti-Corruption Questionnaire

Administration for Children's Services	Department of Transportation
Board of Elections	Department of Veterans' Services
Business Integrity Commission	Department of Youth and Community Development
Civilian Complaint Review Board	Economic Development Corporation
Commission on Human Rights	Financial Information Services Agency
Department for the Aging	Fire Department
Department of Buildings	Landmarks Preservation Commission
Department of City Planning	Law Department
Department of Citywide Administrative Services	Mayor's Office
Department of Correction	Mayor's Office of Contract Services
Department of Consumer and Worker Protection	Mayor's Office of Criminal Justice
Department of Cultural Affairs	New York City Emergency Management

Department of Design and Construction	New York City Employee Retirement System
Department of Environmental Protection	New York City Housing Authority
Department of Finance	New York City Housing Development Corporation (HDC)
Department of Health and Mental Hygiene	Office of Administrative Trials and Hearings
Department of Housing Preservation and Development	Office of the Chief Medical Examiner
Department of Investigation	Office of Technology and Innovation
Department of Parks and Recreation	Office of Labor Relations
Department of Probation	Office of Management and Budget
Department of Records and Information Services	Office of Tax Administrative Appeals
Department of Sanitation	Police Department
Department of Small Business Services	School Construction Authority
Department of Social Services	Taxi and Limousine Commission

B. Database Numbers

The Questionnaire asked agencies to list and describe all databases that they control, either directly or via contract, or for which they are the primary user. The Questionnaire prompted agencies to include the type of data contained within each database, who owns the database, whether the data is captured manually or electronically (or both), and whether the data is stored on-premises or in the cloud. For Questionnaire purposes, DOI defined a "database" as any computer-based system, whether on-premises or in the cloud, that collects and stores data, automatically or manually, that can be used, accessed, or modified by way of a user interface. A database could be proprietary to an agency, to the City, or to an outside vendor, and could be internal or public-facing. To guide agencies in their responses, the Questionnaire set forth a non-exhaustive list of database examples, including case management systems, records management systems, financial management systems, property management systems, learning management systems, human resource management systems, supply chain management systems, and electronic inventories and libraries. The Questionnaire clarified that database did not include email systems or archives; spreadsheets; or third-party databases, such as LexisNexis, used exclusively for conducting research.

In total, agencies reported over 1,200 databases that are controlled by the agency or for which the agency is the primary user. The number of databases per agency varied dramatically—three agencies reported as few as one, while one agency reported as many as 225 databases. Of the agencies which provided the requisite details, roughly over 80 percent of databases were proprietary to, owned, and hosted by the agency, with the remainder operated or owned by third party vendors. Of the agencies which provided the requisite details, roughly 43 percent of databases involved manual entry,[10] 38 percent involved automatic entry, and the remainder incorporated elements of both automatic and manual entry.

C. Data Integrity Measures Per Database

The Questionnaire asked agencies to detail the data integrity measures in place to control the accuracy, completeness, and quality of the data maintained in the five, core databases identified. To guide responses, the Questionnaire provided agencies with a non-exhaustive list of data integrity measures, including user access controls, audit trails, multi-user approval for data changes, disaster recovery plans, and separation of duties. Instructions in the Questionnaire directed agencies to omit cybersecurity measures from their description of data integrity measures. Below are highlights of some of the identified data integrity measures:

Limit Access, User-Based Roles. Thirty of the 48 responding agencies reported that they had data integrity measures limiting access to databases. A number of these agencies explained that they limited database access to those employees who required it to perform their duties. This approach is also referred to as "role-based access control" ("RBAC").[11] RBAC restricts access to digital resources based on a user's role in an organization.[12] RBAC is based on the principle of least privilege, which means that any user's level of access privileges should be calibrated to their specific needs based on their role.[13]

Audit Trails. At least 22 agencies reported using audit trails. An audit trail is a record of computer events involving user activities on an operating system or an application.[14] For data integrity purposes, audit trails track and create a log of activity related to a set of data and typically include some or all of the following information:

- Who viewed, modified, or moved data?

[10] "Manual entry" means directly entering data into a database. When data is entered manually, there are no controls as to the data entered and potentially no audit trail.
[11] David F. Ferraiolo and D. Richard Kuhn, *Role-Based Access Controls*, National Institute of Standards and Technology (1992), https://csrc.nist.gov/files/pubs/conference/1992/10/13/rolebased-access-controls/final/docs/ferraiolo-kuhn-92.pdf.
[12] David F. Ferraiolo and D. Richard Kuhn, *Role-Based Access Controls*, National Institute of Standards and Technology (1992), https://csrc.nist.gov/files/pubs/conference/1992/10/13/rolebased-access-controls/final/docs/ferraiolo-kuhn-92.pdf.
[13] David F. Ferraiolo and D. Richard Kuhn, *Role-Based Access Controls*, National Institute of Standards and Technology (1992), https://csrc.nist.gov/files/pubs/conference/1992/10/13/rolebased-access-controls/final/docs/ferraiolo-kuhn-92.pdf.
[14] 1 Cybersecurity Resilience Planning Handbook § 13.01 (2020).

- When was the data created, modified, relocated or deleted?
- How did a user access the data?
- What was the query used to find and access the data?
- Was the access authorized?
- Were the changes approved by an authorized person?
- Were any access rights abused?

Audit trails are crucial tools to enhance agency accountability because they create records of user activities that can help with detection of data misuse or improper behavior by enabling agencies to identify how and when an action took place, as well as the responsible user.[15] But an audit trail is only as useful as the data it tracks—the more detail they show, the more helpful they can be. Agencies were not specifically asked about audit trails, but most proactively reported that the audit trails used had the capacity to, at a minimum, track user log in details, time stamps of actions, and data changes made by a user.

Periodic Internal Audits. A few agencies also reported conducting periodic internal audits. Periodic internal audits can serve many purposes: they allow agencies to review data contained within the identified databases to search for suspicious or malicious activity as well as to identify incomplete, inaccurate or inconsistent data and provide a better understanding of the actions required to achieve an appropriate level of data quality.[16] By regularly conducting data quality audits, agencies can take immediate action to improve problematic processes or identify users who may benefit from corrective actions such as additional training.

OTI Policies and Procedures. In addition to agency-specific policies and procedures, mayoral agencies are also subject to citywide policies and procedures related to data promulgated by the Office of Technology and Innovation ("OTI"). For instance, all mayoral City agencies must follow strict guidelines around the collection, maintenance, disclosure and use of personal identifying information. Pursuant to OTI policy, City agencies must protect the quality, integrity, and accuracy of identifying information and take reasonable steps to correct, update, or securely dispose of inaccurate or outdated identifying information.[17] The policy also creates a reporting structure for agencies to notify OTI if a breach of the agency's security has occurred where an individual's identifying information has been or is

[15] 1 Cybersecurity Resilience Planning Handbook § 13.01 (2020).

[16] Ed Gelbstein, Ph.D., *IS Audit Basics: The Domains of Data ad Information Audits*, ISACA (Dec. 2, 2016), https://www.isaca.org/resources/isaca-journal/issues/2016/volume-6/is-audit-basics-the-domains-of-data-and-information-audits.

[17] NYCOTI, *Citywide Privacy Protection Policies and Protocols* (Feb. 6, 2023), available at: https://www.nyc.gov/assets/oti/downloads/pdf/citywide-privacy-protection-policies-protocols.pdf.

reasonably believed to have been accessed, acquired, disclosed, or used without authorization.[18] Another OTI policy governs critical data and removal of data devices (such as USB drives, CDs, external drives, etc.) and requires that they must be protected by appropriate physical means from modification, theft, or unauthorized access and must meet the requirements set forth in the Citywide Portable Data Security Policy.[19] And finally, OTI policy acknowledges that "[p]rotecting data's confidentiality, integrity and availability is a principle that must be maintained at all times."[20] Methods provided to advance this principle include, but are not limited to, data encryption and application access based on individual identification and authentication.[21]

Creative Data Integrity Measures. One agency has reported that it initiated a multi-year agency-wide data governance initiative to focus on a structured approach to data governance. This agency reported that the initiative is overseen by senior leadership and a Data Governance Advisory Board—a working group that includes representatives from different agency units. According to the agency, the goal of the initiative is to provide structure and guidance necessary to use data and information effectively and efficiently. Major areas of focus will include data discovery, data accessibility, data accountability, and data integrity. The data integrity focus will target data accuracy and quality, data retention, and data security. As part of this initiative, the agency also created and filled the role of Director of Data Management.

Another agency reported that, as it was preparing to migrate operations from a legacy system to an updated one, it created a "Data Cleansing Committee" responsible for reviewing and resolving any data discrepancies to ensure the agency had quality data before migrating systems. The agency also reported having a "Business Rules and Data Division" responsible for writing business rules for data maintenance and identifying areas for improvement and a robust "Information Security and Privacy Policy" which acknowledged that the agency processed a high volume of information that needs to be protected wherever it is contained. That policy, unique to that agency, classified data as Highly Confidential / High Risk, Confidential, or Low risk and each category invoked rules specific to its classification.

[18] NYCOTI, *Citywide Privacy Protection Policies and Protocols* (Feb. 6, 2023), available at: https://www.nyc.gov/assets/oti/downloads/pdf/citywide-privacy-protection-policies-protocols.pdf.

[19] NYCOTI, *Portable Data Security Policy* (Sept. 9, 2014), available at: https://www.nyc.gov/assets/oti/downloads/pdf/vendor-resources/portable-data.pdf. *See also* NYCOTI, *User Responsibilities Policy* (Sept. 9, 2014), available at: https://www.nyc.gov/assets/oti/downloads/pdf/vendor-resources/user-responsibilities.pdf.

[20] NYCOTI, *Citywide Application Security* (Nov. 1, 2018), available at: https://www.nyc.gov/assets/oti/downloads/pdf/vendor-resources/citywide-application-security-p-as-01.pdf.

[21] NYCOTI, *Citywide Application Security* (Nov. 1, 2018), available at: https://www.nyc.gov/assets/oti/downloads/pdf/vendor-resources/citywide-application-security-p-as-01.pdf.

D. Data Integrity Risks

For each of the five databases identified by agencies, the Questionnaire asked agencies to detail any data integrity risks not addressed by the data integrity measures they have in place.

Ten agencies reported that they do not utilize some of the data integrity measures discussed above. For example, at least six agencies allow database access and permissions to all employees, or at least to more employees than is necessary based on roles. One agency even reported that a handful of employees shared a single login to a database, which poses security and recovery risk as well as undermines the purpose of audit trails. At least three agencies responded that they were not conducting or tracking audit trails at all or could improve their use of audit trails.

Several agencies identified staffing and retention issues as data integrity risks.[22] Agencies explained that rapid agency turnover is resulting in diminished database user education. One agency reported: "The primary risk lies in the potential turnover of the personnel responsible for maintaining this application and its documentation." Another echoed this concern and described a key risk as "legacy systems using technology that has become progressively harder to continue supporting given the diminished pool of experienced resources and budgetary constraints that prohibit us from hiring." And a third reported that main databases have "recurring resource constraints due to staff shortages so that often the production support and development team have difficulty fulfilling the roles with different staff."

Age of platform and third-party vendors were identified by at least three agencies as data integrity risks. Both concepts were intertwined, according to responses, as outdated databases tended to involve outside vendors who failed to update and enhance the platforms. As one agency put it: "The System is almost 20 years old, and the vendor may not be dedicating its most robust resources to its maintenance." That same agency noted its efforts, already underway, to replace that system with a new product and implementation partner. A separate agency echoed the same concern and remediation attempts: "System and hardware are very old. Currently looking to upgrade/replace the entire system."

Along these same lines, some agencies reported inherent risks in using third-party vendors. One agency reported: "While [the System] maintains strict data integrity rules and security protocols, it is a third-party application. There are

[22] For more information about the City's human capital management challenges and their impact on anti-corruption activities, please see DOI's 2022 Annual Anti-Corruption Report, available at: https://www.nyc.gov/assets/doi/press-releases/2023/May/19AntiCorrRpt.Release.05.01.2023.docx.pdf.

inherent risks associated with this: [s]toring sensitive data on third-party servers creates a risk of data security breaches or unauthorized access." This may be an unavoidable risk, as some agencies do not have the capacity or staffing to maintain their own databases. In that event, it is important for agencies to carefully vet and scrutinize possible data system vendors.

Finally, over half of the responding agencies reported that there were no data integrity risks with respect to one or all of their five key databases. A figure this high suggests that agencies may not be capturing or considering the variety of ways that data integrity may be undermined or compromised.

E. Written Data Integrity Policies or Procedures

The Questionnaire asked agencies whether they have any written policies or procedures related to data integrity. While an earlier prompt asked agencies to identify which data integrity measures they had in place, this question specifically probed whether those measures—or other policies or procedures related to data integrity—were documented in a formal, written policy. Those agencies who responded in the affirmative were also asked to send a copy of those written policies and procedures to DOI.

Of the 48 agencies that responded to the Questionnaire, 18 reported that they had no written policies on data integrity. Each of the 18 agencies reported in other Questionnaire responses that they had practices in place to address data integrity. Some of the agencies lacking written policies routinely deal with highly sensitive data.

Thirty agencies indicated that they had written polices or procedures in place. Notably, though, at least five of those agencies submitted OTI's citywide policies related to data and appeared to lack agency-specific policies or procedures. And even within the policies and procedures submitted that were agency-specific, at least three dealt exclusively with protection from unauthorized third parties and thus would more accurately be described as cybersecurity policies rather than data integrity policies. This indicates that some agency policies do not reflect the distinction between cybersecurity—threats from external third-party actors—and data integrity—ensuring the accuracy and consistency of data.

Further, a few agencies submitted audit policies that did not mention or address data and instead outlined general audit mechanisms. For instance, one agency submitted an external business audit guide which included procedures for staff to follow when implementing suggestions received from third party audits of the agency's operations and functions. While some principles of a successful audit from this policy may be transferrable to a data audit, this policy cannot be fairly characterized as specifically related to data integrity.

Below are some examples of agency-specific written data integrity policy and procedures that DOI received in response to the Questionnaire. The below could be instructive to those agencies that lack written policies but wish to implement them:

- "User Access Management" and "User Access Management Procedure": These data oversight policies and procedures govern the creation, modification, and deletion of users' logical and physical access by, among other things, ensuring former employees or those who are out on leave do not retain data access.
- "Access Control Policy": This access review policy defines the requirements for secure access to agency computer and communications systems including access restriction such as the revocation of data access to inactive and disabled accounts.
- "Operating Procedure": This policy sets out steps to investigate active and closed cases suspected of fraud and abuse based on computer match data and allegations received from the public and other governmental agencies.
- "Data Integrity QM Policy": This policy establishes and explains the mandatory process referred to as Quality Management ("QM") to ensure data integrity.
- "Data Governance Policy, Standard, and Process": This policy discusses the mandate and scope of the Field Audit Review Unit, a unit responsible for quality assurance of data entry.
- "IT Access Management Least Privilege": This policy dictates that least privilege users or resources will be provided with the minimum privileges necessary to fulfill their roles and responsibilities.

Responses to this question illustrated different approaches taken by City agencies with respect to data integrity policies and procedures, including that some have no policies at all.

F. Use of Data to Identify, Prevent, Reduce, or Eliminate Instances of or Risks of Corruption, Fraud, Waste, and Abuse

The Questionnaire asked agencies whether the data contained in any of the identified databases is used to identify, prevent, reduce, or eliminate instances or risks of corruption, fraud, waste, or abuse. To guide responses, DOI provided a non-exhaustive list of ways in which data could be used in this way, including reviewing contractor performance, inventory control, and auditing of invoices and billing. Fifteen agencies indicated that they do not use their data for the purposes of identifying, preventing, reducing, or eliminating instances or risks or corruption, fraud, waste, or abuse.

Thirty-three agencies responded that they do use data for this purpose and reported a variety of different methods. For instance, one agency utilizes an application to audit inventory and billing, and even has a special monitoring unit responsible for flagging inconsistencies in data to prevent, mitigate, or eliminate instances of corruption, fraud, waste, and abuse. Other agencies have similar methods: one agency reported that it uses its data to cross-reference invoice accuracy, while a third responded that it uses its data to cross-reference and audit stockroom inventory.

G. Data Monitoring Staff or Units

The Questionnaire asked agencies whether they have staff or units responsible for analyzing, monitoring, and/or auditing data contained in any of the identified databases. If so, agencies were asked to identify the budgeted headcount, what expertise that staff possesses, what types of analyses and audits were conducted and how often, what reports or work product is generated from such analyses, and whether the agency has any written policies or procedures pertaining to such analyses.

Eight agencies reported having no staff or units responsible for analyzing, monitoring, and/or auditing data contained in any of the identified databases. Forty agencies reported having staff or units for these purposes.

Of the 40 agencies with such staff, the budgeted headcount of those groups varied, but appeared to be largely proportional to the size of the particular agency. For instance, larger agencies reported having as many as 72 staff members and smaller agencies had as few as one or two staff members. The experience of those staff members also varied. Agencies reported a mix of advanced degrees, civil service exams or certifications, and general expertise in the data system itself or experience in City government. Most staff or units reportedly generate excel spreadsheet reports for review and analysis, while some produce formal audit reports based on the review of data. Many agencies reported using this work product to correct or amend their data, to make policy recommendations to superiors or to other state or federal agencies, and to make assessments of data integrity and compliance with applicable policies or rules.

Several years ago, DOI established a stand-alone Data Analytics Unit and their work has shown what DOI can accomplish with access to City data.[23] DOI is currently seeking to expand its capacities to analyze data and engage with analytics staff at other agencies in connection with City investigations and oversight. Specifically, DOI is in the early stages of an effort to expand DOI's direct access to City agency databases, which will enable DOI to more efficiently access data, more effectively assemble useful data sets, and more thoroughly analyze data to identify potential corruption, fraud, waste, and abuse. The success of this initiative depends upon the cooperation of City agencies.

H. Algorithmic Tools or Artificial Intelligence

Lastly, the Questionnaire asked whether agencies use algorithmic tools or artificial intelligence ("AI") to analyze, monitor, or audit data to prevent corruption, fraud, waste, or abuse. Only six agencies indicated that they use AI in this way.

Those who did use AI shared some of the ways that the technology has aided them in their efforts to identify corruption, fraud, waste, and abuse. For instance, one agency identified a tool with algorithms that check index data against set conditions and flag submissions as potentially fraudulent based on the presence of certain conditions. That same agency identified a second tool which uses two commercially available software suites to provide a fraud risk score for credit card authorization requests; both programs rely in part on algorithm-based evaluations of transaction characteristics. A second agency evaluates log-in data to flag and track off-premises VPN access to limit off-premises access to malicious internet content and unauthorized agency websites.

The Questionnaire then asked whether any agencies not currently using algorithmic tools or AI for anti-corruption purposes were discussing or considering

[23] For instance, in January of 2024, DOI investigated and substantiated claims of artificial data manipulation of the PATH's publicly-reported Monthly Eligibility Rate by unnecessarily delaying DHS's final determination that families had been deemed eligible for shelter. *DOI Issue Report on Disclosure of Overnight Stays at the PATH Intake Center in Summer 2022 and the Manipulation of the Publicly-Reported PATH Eligibility Rate From 2017 to Early to Mid-2022*, available at: https://www.nyc.gov/assets/doi/press-releases/2024/January/02DSSRelease.Rpt.01.09.2024.pdf. DOI's investigation included a review of PATH data to conclude publicly-reported data was being manipulated in this way. *Id.* In January of 2022, DOI utilized NYCERS' data to investigate and charge a Bronx resident with stealing more than $50,000 in City pension funds issued to a deceased pensioner. *DOI Arrests Bronx Resident On a Charge of Stealing More Than $50,000 in City Pension Funds*, (Jan. 11, 2022), https://www.nyc.gov/assets/doi/press-releases/2022/January/01Pension_01112022.pdf. There, DOI began investigating after examining NYCERS' records to identify all potentially deceased pensioners, which revealed this specific scheme which took place from November 2015 to May 2018. *Id.*

doing so. Only three of the forty agencies indicated that they were engaged in such discussion or consideration. Several factors may potentially explain the minimal use or consideration of algorithmic tools or AI for anti-corruption purposes. Agencies may be hesitant to use an emerging technology and prefer to await further developments before incorporating such technological tools into their anti-corruption efforts. Moreover, non-IT agency staff may not be familiar with how this technology works and, as a result, may have concerns with using technology platforms with which they are unfamiliar. Concerns may also involve privacy and the use of personal identifying information within AI tools or platforms. Mitigation could involve contracts and confidentiality agreements with AI providers to avoid risks of exposure of City data.

Agency responses to this Questionnaire prompt revealed that most City agencies have not yet pursued or considered potential opportunities to utilize algorithmic tools or AI for anti-corruption purposes. However, in October 2023, Mayor Eric Adams and Chief Technology Officer Matthew Fraser released the Adams administration's comprehensive "New York City Artificial Intelligence Action Plan" (the "Plan") to develop a framework for City agencies to evaluate AI tools and risks.[24] The Plan outlined 37 key actions, including, among other things, establishing a framework for AI governance, creating an external advisory network to consult with stakeholders across sectors, and building AI knowledge and skills in City government to prepare City employees to effectively and responsibly work with and on AI.[25] Similar action is taking place at a state level.[26]

IV. Recommendations and Conclusion

Understanding and addressing data integrity risks is crucial for City agencies to ensure the accuracy of the data they use for their work. Additionally, City data provides valuable information for use in City efforts to combat risks of corruption, fraud, and abuse. As discussed above, DOI is in the early stages of an effort to expand its direct access to City agency databases, so that DOI can more efficiently and effectively use City data to perform its oversight role and protect the City from

[24] *Mayor Adams Releases First-of-Its-Kind Plan For Responsible Artificial Intelligence Use In NYC Government*, NYC (Oct. 16, 2023), https://www.nyc.gov/office-of-the-mayor/news/777-23/mayor-adams-releases-first-of-its-kind-plan-responsible-artificial-intelligence-use-nyc#/0.

[25] Available at: https://www.nyc.gov/assets/oti/downloads/pdf/reports/artificial-intelligence-action-plan.pdf.

[26] In the 2024 State of the State address, Governor Hochul announced a proposal for New York to invest $400 million to create and launch an AI computing center in upstate New York to promote responsible research and development and determine AI use cases that can benefit the public. *Available at*: https://www.governor.ny.gov/sites/default/files/2024-01/2024-SOTS-Book-Online.pdf. At the same time, the New York state Office of Information Technology Services issued a new policy to establish guidelines governing the evaluation and adoption of AI systems by state agencies. *Available at*: https://its.ny.gov/system/files/documents/2024/01/nys-p24-001-acceptable-use-of-artificial-intelligence-technologies-_1.pdf.

corruption risks. A critical mechanism for enhancing the City's use of data for anti-corruption purposes will be providing DOI with access to relevant databases. DOI will require the cooperation of City agencies in facilitating such database access to advance the shared objectives of DOI and other City agencies in preventing and identifying corruption.

This Anti-Corruption Report set forth several key findings related to data integrity at City agencies:

- The majority of City agencies utilize some combination of data integrity best practices, including user-based limitations on access, audit trails, and periodic internal audits.

- Agencies self-reported key risks to data integrity. Those risks included lack of role-based permissions and access, increased employee turnover and retention issues, as well as the advanced age of certain database platforms and maintenance of those platforms by third-party, outside vendors.

- Thirty agencies indicated that they had written policies or procedures in place governing data integrity, though five of those agencies submitted OTI's citywide policies and did not submit any agency-specific policies on data integrity. A few of those agencies submitted audit policies that did not explicitly mention or address data. Eighteen agencies reported they had no written policies on data integrity, though each of those agencies reported having practices in place to address data integrity—presumably, then, not memorialized in writing.

- Thirty-three agencies responded that they do use data to identify, prevent, reduce, or eliminate instances or risks of corruption, fraud, waste, or abuse. Agencies reported methods including cross-referencing data against other, set inputs to flag issues.

- Only six of forty-eight agencies responded that they use AI to analyze, monitor, or audit data to prevent corruption, fraud, waste, or abuse.

- Eight agencies reported having no staff or units responsible for analyzing, monitoring, and/or auditing data contained in any of the identified databases. Of the forty agencies with staff, headcounts varied but appeared largely proportional to the size of the particular agency. The experience of those staff members also varied and included a mix of advanced degrees, civil service exams or certifications, and general expertise in the data system itself or experience in City government.

Based on these findings, DOI recommends that City agencies assess their current data integrity policies and practices to evaluate whether they adequately promote data integrity and sufficiently utilize data to address risks of fraud, corruption, and abuse. As part of that assessment, agencies should consider the feasibility and applicability of these safeguards in light of their specific needs:

- Ensure that the agency has a written data policy that includes provisions regarding data governance, such as access control and disaster recovery procedures. Such data policy should be periodically reviewed and updated as necessary.

- Appoint a data officer responsible for setting the data policy, determining access, and reviewing compliance.

- Where possible, phase out the manual entry of data, moving to electronic input only. With respect to data deletion, limit deletion authority to a small universe of appropriate supervisory staff.

- Control database access based on roles or groups to which individuals are assigned so that access is consistent across similarly situated staff.

- Test and simulate disaster recovery procedures periodically to ensure that they will work as intended when actually needed.