Office of Internal Audit

Performance Audit 18-04: Open Data

December 2018

City Auditor
Stan Sewell, CPA, CGFM, CFE

Senior Auditor
Jeff Connor, JD, CFE

OFFICE OF INTERNAL AUDIT

Stan Sewell, City Auditor

December 28, 2018

To: Mayor Andy Berke
City Council Members

Subject: Open Data Audit (Report #18-04)

Dear Mayor Berke and City Council Members:

The attached report contains the results of our audit of the Open Data program. Our audit found that the Office of Performance Management and Open Data has developed an efficient and sustainable open data program for the City of Chattanooga that promotes informed decision-making, transparency, and robust community engagement. However, we also determined that certain requirements for the open data program established by Executive Order 2014-04 have not been fully implemented. In order to address the noted areas for improvement, we recommended actions to: 1) remediate areas of non-compliance with the City's open data initiatives; 2) develop a comprehensive policy and procedural framework for the Open Data program; and 3) strengthen privacy protections for information published on the open data portal.

We thank the management and staff of the Office of Performance Management and Open Data, Transportation, Public Works, and Police departments for their cooperation and assistance during this audit.

Sincerely,

Digitally signed by Stan Sewell
DN: cn=Stan Sewell, o=City of Chattanooga, ou=Office of Internal Audit, email=ssewell@chattanooga.gov, c=US
Date: 2018.12.28 09:22:31 -05'00'

Stan Sewell, CPA, CGFM, CFE
City Auditor

Attachment

cc: Audit Committee Members
Stacy Richardson, Chief of Staff
Maura Sullivan, Chief Operating Officer
Tim Moreland, Director, Office of Performance Management and Open Data
Jim Arnette, Tennessee Local Government Audit jim.arnette@cot.tn.gov

1001 Lindsay Street • Chattanooga, Tennessee 37402
(423) 643-6200 • FAX: (423) 643-6204 • E-MAIL: ssewell@chattanooga.gov

TABLE OF CONTENTS

AUDIT PURPOSE

This audit was conducted in accordance with the Office of Internal Audit's 2018 Audit Agenda. The objectives of this audit were to determine if:

- City departments and agencies have complied with Executive Order 2014-04 establishing an Open Data program.
- The Office of Performance Management and Open Data maintains a comprehensive open data policy and procedures to safeguard against the publication of inaccurate data.
- The Open Data program has a reliable process for protecting sensitive data and personally identifiable information from inadvertent public disclosure.

BACKGROUND

Data is a valuable resource and a strategic asset to the City of Chattanooga, its partners, and the public. Managing data as an asset and making it available, discoverable, and usable not only strengthens public accountability and promotes efficiency and effectiveness in government, but also has the potential to create economic opportunity and improve citizens' quality of life.[1]

To improve the transparency of government functions and make information resources more accessible to the public, Mayor Andy Berke issued Executive Order 2014-04 (the "Mayor's Order") establishing an open data program for the City of Chattanooga in May 2014. The Mayor's Order requires each city agency and department to identify and publish information open to public inspection under the Tennessee Public Records Act.[2] The Mayor's Order further directs city agencies to treat data as open by default for publication on Chattanooga's open data portal.[3]

The City's open data is published on a single web portal maintained by the Chattanooga Public Library. All open data is required to be catalogued in a public manner that clearly articulates: 1) if the data is available in digital form; 2) if the data is machine-readable; 3) the

[1] Project Open Data, OMB Memorandum M-13-13, *Open Data Policy—Managing Information as an Asset* (May 19, 2013)

[2] Tenn. Code §10-7-503, *et seq.*

[3] The federal government and over 85 cities across the nation, including Nashville, Knoxville and Memphis, have open data schemas similar to Chattanooga's program.

agency responsible for the data; 4) if the data and metadata[4] are clearly documented; and 5) if the data is available on the open data portal, and if not, the reason(s) why the data is unavailable.

The Office of Performance Management and Open Data administers the City's open data program, which is purposely designed to increase the availability of public information and leverage data resources to improve the efficacy and performance of city government functions.

Exhibit 1. Office of Performance Management and Open Data Metrics

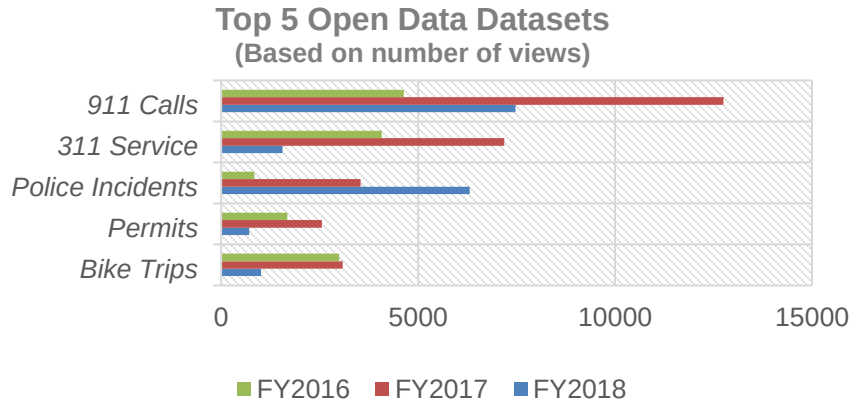
Calendar Year	2016	2017	2018
Operating Budget (1)	$208,968	$211,803	$218,772
Published Datasets	202	269	283
Data Pages Viewed	201,060	68,501	58,357
Website Embeds	1,106	3,021	5,611
Data Rows Created	95.1M	37.6M	3.2M

(1) Reported on a fiscal year basis in the CABR

Source: FY2018 Open Data Annual Report

The Mayor's Order also establishes an Open Data Advisory Group whose members include data coordinators from each city agency. The advisory group is responsible for assisting city departments in the fulfillment of their open data responsibilities and developing new processes to promote online public access to city datasets.

Exhibit 2. Open Data Usage in Chattanooga



Source: FY 2018 Open Data Annual Report

[4] The National Information Standards Organization (NISO) defines metadata as: "*data about data [including] information created, stored and shared to describe data… [Metadata] informs our understanding of what information is, how it relates to other things, and how it can be used*".

FINDINGS AND RECOMMENDATIONS

Remediate Areas of Noncompliance with the Mayor's Executive Order

The Office of Performance Management and Open Data has made tremendous progress towards fulfilling the Mayor's mandate for a robust open data initiative. However, some city departments and agencies have not fully complied with the Mayor's Order.

Executive Order 2014-04 establishes specific goals and expectations for the City's open data program, including:

- Promote and prioritize dataset availability;
- Strengthen open data governance;
- Define roles and responsibilities for open data personnel;
- Develop a comprehensive inventory of City datasets;
- Create accountability and compliance functions; and
- Establish guidelines for technology-related procurements.

Based on the Office of Internal Audit's assessment and findings, we determined the following requirements established by the Mayor's Order for the open data program have not been fully implemented:

Dataset availability

a) Not all city agencies have published their existing priority datasets on the open data portal. For example, the Finance department has several high-priority datasets that remain unpublished without providing clear rationalization for the delay in availability on the open data portal.
b) Datasets are not prioritized for publication based on the frequency in which they are requested by the public through the Tennessee Public Records Act.
c) There is a lack of supporting documentation verifying that city agencies are treating newly-created datasets as open by default.

Data Inventory

d) Numerous datasets on the open data inventory provide no explanation as to whether the required metadata is clearly documented.
e) The data inventory does not clearly define the reason(s) why certain datasets remain unpublished. Similarly, the reason(s) given for why some datasets have not been published is not clearly articulated.

Technology-Related Procurements

f) Contract guidelines have not been developed to promote open data policies in technology-related procurements.
g) Guidelines have not been established to promote the City's open

data policies, including, where appropriate, requirements to post data on the open data portal or to make data available through other means.

Recommendation 1:

We recommend the Office of Performance Management and Open Data continue to work with the city agencies and departments to address areas of noncompliance with the Mayor's Order. We further recommend the Office of Performance Management and Open Data, in conjunction with open data coordinators from each city agency, perform an annual review of the open data inventory to promote the availability of priority datasets and ensure the published datasets and metadata are accurate, reliable, and complete.

Auditee Response: *We concur with the audit findings and recommendation. We will update our open data inventory to provide the reasons behind the delay of publishing priority open datasets on the open data portal, and include more information on the metadata and why datasets are not yet published.*

Additionally, we will work with the City's public records coordinator to ensure high frequency datasets are published on the open data portal. In the past, we have had trouble getting the raw data on requests. In addition to difficulty getting the data, most requests are in free form text which makes it difficult to categorize and analyze. We have had additional conversations with the public record coordinator about getting access to the raw data, and have begun discussions on how we might be able to pull actionable insights from the free form text.

Given our limited staffing we have not been able to aggressively enact the open by default policy called for in the Mayor's Order. Instead, we have focused our efforts on opening datasets shown to be a priority through our internal data inventory, top datasets listed in the U.S. open data census, and by looking at the most popular open datasets in other cities' open data programs.

We will continue to coordinate with the Department of Information and Technology (DIT) to develop guidelines to promote open data policies in technology-related procurements. We prefer to automate our open datasets for long-term sustainability instead of one-off publishing of open data by the agency. Most departments do not have the technical expertise to automate the publishing of their datasets to the open data portal and rely on the help of the Office of Performance and Open Data. This is accomplished through an annual update to the City data inventory each year.

Develop and Implement a Comprehensive Open Data Policy and Procedures

Aside from the Mayor's Order, the Office of Performance Management and Open Data does not have a formal written policy to ensure data received from the various city agencies is accurate and complete. Our analysis of best practices from other government open data programs concludes that a comprehensive policy and procedural framework is essential to institutionalizing the principles of effective information management at each stage of the information lifecycle.

In addition to evaluating management controls over open data processes and procedures, we sample tested select datasets on the open data portal for accuracy and completeness. We obtained dataset samples from the city agencies whose datasets were most often accessed by the public from FY2017-2018 (Bike Trip Rentals, 311 Service Requests, and Police Incidents)[5], and traced the sample datasets to the information published on the open data portal. Our findings for each dataset are summarized below.

Bike Trip Rentals

a) Bike trip rental transactions from the Comet system[6] displayed Start Times/End Times that did not match the data contained in the open data portal. In each transaction we tested, the Start Times/End Times listed in the open data portal were delayed by four (4) hours.
b) In some instances, the four (4) hour Start Time/End Time variance resulted in incorrect transaction dates on the open data portal.

311 Service Requests

c) Fifteen (15) percent of the 311 transactions we examined from the Accela system had no corresponding record/data in the open data portal.
d) The Accela 311 data displayed time stamps in the Created Date field that did not match the time stamps for the data on the open data portal. In each transaction we tested, the time stamp displayed on the open data portal was delayed several hours.
e) The 311 transactions on the open data portal did not contain data fields or metadata for the Title, Request Type, Due Date, or Completion Date—fields included in the Accela system. In many instances, the Description field on the open data portal was

[5] Hamilton County Emergency 911 data was not selected as part of our test sample because the datasets are no longer updated to the open data portal.

[6] The Comet system is a secure cloud-based application maintained by Chattanooga's vendor, Shift Transit. All bike trip transactional data is collected by Shift Transit and stored electronically on the Comet system.

blank making it impossible to determine the nature of the service request.

f) The Request Type Code field on the open data portal did not match any corresponding data field in the Accela system.

Police Incidents

g) Forty-seven (47) percent of the police incident data we tested from the open data portal had no corresponding record/incident in the Chattanooga Police Record Management System (RMS).

h) Twelve (12) percent of the police incidents we examined from the open data portal contained information in the Incident Description field that did not match the corresponding data in the Incident Description field in the RMS system.

Recommendation 2:

We recommend the Office of Performance Management and Open Data establish controls to ensure datasets on the open data portal are accurate, reliable and complete. We further recommend that the Office of Performance Management and Open Data develop a comprehensive open data policy and procedures that define the principles governing the program and describe, in precise terms, the expectations and accountabilities for department participation in the program.

The policy should require city agencies to collect and create information in a way that supports downstream information processing and dissemination, including machine-readable and open formats, data standards, and common core and extensible metadata for all new information. It should also require city agencies to ensure information stewardship through the use of open licenses and review of information for privacy, confidentiality, security, or other restrictions to release.

Lastly, the policy should require city agencies to build and modernize information systems in a way that maximizes interoperability and information accessibility, maintains data asset inventories, enhances information safeguards, and clarifies information management.

Auditee Response: *We concur with the audit findings and recommendation. We currently have several open data program support tools and resources available, including an open data workflow and open data coordinator dataset review checklist, which we encourage all city agencies and departments to use to ensure data accuracy and completeness.*

We were not aware of the Start Time/End Time inconsistencies in the datasets identified by the audit, and are looking into root causes and

potential fixes. We will also be reviewing other datasets to make sure this is not a broader issue with how we are handling dates and times.

With regard to police incident data, we believe the reason some of the incident data on the open data portal has no corresponding record in the RMS system is due to the fact that police incident data maintained by the Chattanooga Police Department (CPD) does not contain "miscellaneous" incidents. Since it is such a large volume of incidents we thought it would be good to include them on the open data portal. We are currently working to get more detailed codes from the CPD so we can provide more detailed descriptions on the miscellaneous incidents.

Additionally, we believe the reason some Incident Descriptions fields on the open data portal do not match the RMS fields is because we replaced the incident field with hard coded values based on the unified crime reporting codes. This was done because the description field is free text that officers input and there were concerns that in some rare cases personal information could be included in the description and therefore released on the open data portal.

Strengthen Privacy Protections for Open Data

Although Chattanooga's open data program utilizes a variety of tools to safeguard sensitive data and personally identifiable information (PII), current operational controls do not cover all relevant aspects of a mature disclosure control program.

Importantly, we found no privacy violations or deviations in the datasets we tested on the open data portal. However, our testing revealed inconsistent applications of data de-identification techniques among various datasets on the open data portal. For example, while the Police Incident dataset generalizes incident addresses, Fire Incident and Emergency 911 Call datasets report the precise address. The inconsistent treatment of location data and other potentially identifiable data fields raises the risk of re-identification[7] for all datasets on the open data portal.

Furthermore, the open data program does not currently have formalized procedures to further mitigate potential re-identification risks, such as triggering reevaluation of datasets in light of changing circumstances, or removing datasets if or when re-identification risks rise too high.

---

[7] A key risk of open data is the possibility of re-identification. Even when names and potentially identifying attributes are removed from a dataset, there is a significant risk someone might be able to deduce that the data relates to a specific individual. The risk of re-identification not only exposes data about the individual that would otherwise not be available to the public, but could potentially result in embarrassing, damaging, or life-threatening implications.

Finally, given the breadth of Tennessee's Public Records Law, the open data program is considerably constrained in its efforts to reduce re-identification risks. Because the public records law mandates the disclosure of even personally identifiable information in many circumstances of legitimate public interest, the open data program must be especially cautious about releasing de-identified records that may be "unlocked" or re-identified through other information subject to public records requests.

Recommendation 3:

We recommend the Office of Performance Management and Open Data consider implementing the following risk mitigation strategies to strengthen existing privacy protections for open data:

a) Develop open data procedures for conducting ongoing screening of approved datasets, elevating the review of risky or sensitive datasets, and remove or modify existing datasets that pose an inappropriate risk of re-identification.
b) Develop or obtain appropriate tools to de-identify unstructured or dynamic data.
c) Continue to deepen workforce privacy training and education efforts.
d) Consult statistical disclosure control experts and invest in programmatic tools to evaluate re-identification risks across all datasets.
e) Perform annual risk assessments of the content available on the open data portal. The outcome of this review should be shared with the city departments and agencies who will help implement risk mitigation strategies.
f) Develop an incident response plan to handle potential privacy breaches or inadvertent disclosures of PII.

Auditee Response: *We concur with the audit findings and recommendations. We are in the process of updating the open data policy to require more detailed privacy standards for the open data program as supporting documentation but not directly written into the policy.*

We currently use Safe FME which allows us to de-identify data, including unstructured data. Due to the difficulty of fully reviewing the de-identification of unstructured data we have often chosen not to include these fields given the risk of potential re-identification. Limited resources (staffing and funding) are major limitations in doing this type of more detailed risk analysis on the open datasets.

We agree that we need a standard operating procedure for potential privacy breaches. Currently, our approach when we are made aware of potential disclosure of inappropriate PII is to immediately make the dataset private, so it is no longer viewable or discoverable by the public, and then to contact the department who owns the data and the public records coordinator. This could be codified in our standards, in much the same way as the open data review standard existing for the departmental open data coordinator.

APPENDIX A: SCOPE, METHODOLOGY AND STANDARDS

Based on the work performed during the preliminary survey and risk assessment, the audit covers the open data program operations from July 1, 2016 to June 30, 2018. When appropriate, the scope was expanded to meet the audit objectives. Source documentation was obtained from City databases and electronic records from external vendors. Original records as well as copies were used as evidence and verified through physical examination.

To develop our recommendations, we interviewed key open data personnel and conducted a detailed analysis of the regulatory criteria and best practices for open data management. Specifically, we analyzed the federal government's framework for open data governance published by the U.S. Office of Management and Budget (OMB), as well as policies and procedures from other well-established open data programs—in particular the City of Seattle, which has achieved national certification from Bloomberg Philanthropies.[8] Using these open data schemas as comparable benchmarks, we analyzed the City's open data program to identify any substantive gaps, deficiencies, and/or areas for improvement.

The sample size and selection were statistically generated using a desired confidence level of 90 percent, expected error rate of 5 percent, and a desired precision of 5 percent. Statistical sampling was used in order to infer the conclusions of test work performed on a sample to the population from which it was drawn and to obtain estimates of sampling error involved. When appropriate, judgmental sampling was used to improve the overall efficiency of the audit.

We relied on computer-processed data contained in the Oracle, Accela and the RMS systems, and assessed the reliability of the data contained in those systems by sufficiently testing the data. Based on our testing, we concluded the data was sufficiently reliable.

We conducted this performance audit from August 2018 to December 2018 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

[8] Bloomberg Philanthropies (What Works Cities) is a nationally recognized organization that evaluates municipal open data programs against a number of standardized criteria to help improve government transparency and accountability.

City of Chattanooga Fraud, Waste, and Abuse Hotline

Internal Audit's Fraud, Waste, and Abuse Hotline gives employees and citizens an avenue to report misconduct, waste or misuse of resources in any City facility or department.

Internal Audit contracts with a hotline vendor, The Network, to provide and maintain the reporting system. The third party system allows for anonymous reports. All reports are taken seriously and responded to in a timely manner. Reports to the hotline serve the public interest and assist the Office of Internal Audit in meeting high standards of public accountability.

To make a report, call 1-877-338-4452 or visit our website:
www.chattanooga.gov/internal-audit