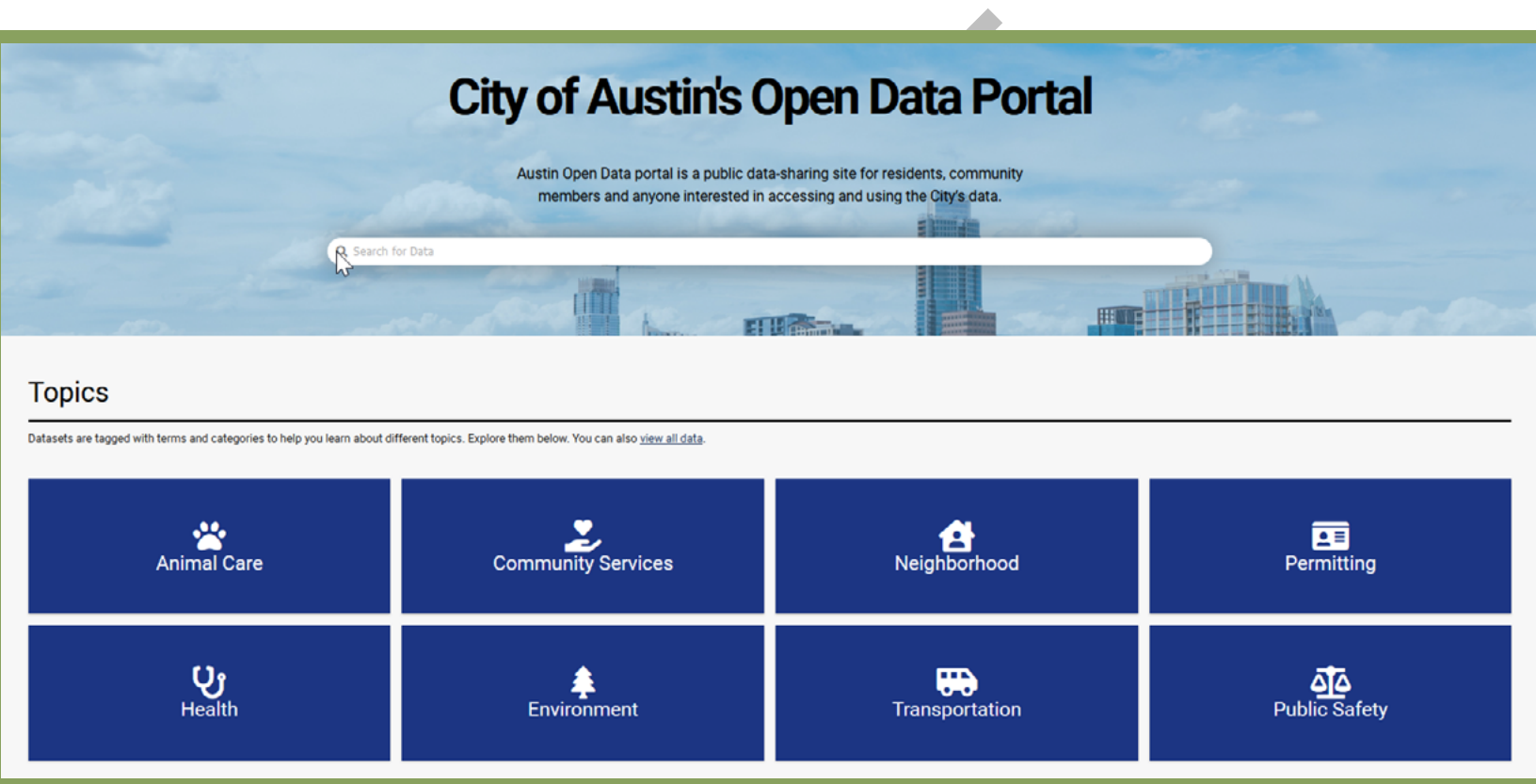


## Audit Report

# Data Reliability: The City provides data through its Open Data Portal, but lacks strategy and oversight

July 2023



City leaders rely on good data to make informed decisions. Similarly, the public may look to data to understand City progress or to hold the City accountable for programs or policies. Unreliable data is a widespread problem for the City. Most of the Office of the City Auditor's recent audits have found issues involving the City's data, many of which remain unresolved. In this audit, we found that data on the City's Open Data Portal are not user friendly or reliable. The City does not have a strategy for displaying its data or ensuring its accuracy. This audit recommends the City centralize oversight of its Open Data Portal. We also recommend creating a Citywide strategy for sharing and presenting our data.

# Contents

Objective and Background .....	2
What We Found .....	4
Recommendations and Management Response .....	11
Scope and Methodology .....	13

Cover: Open Data Portal landing page, <https://data.austintexas.gov/>

## Objective

The objectives of this audit were:

1. Have data reliability issues encountered in previous audits been addressed?
2. Is the information shared by the City on the Open Data Portal reliable, accessible, and useful?

## Background

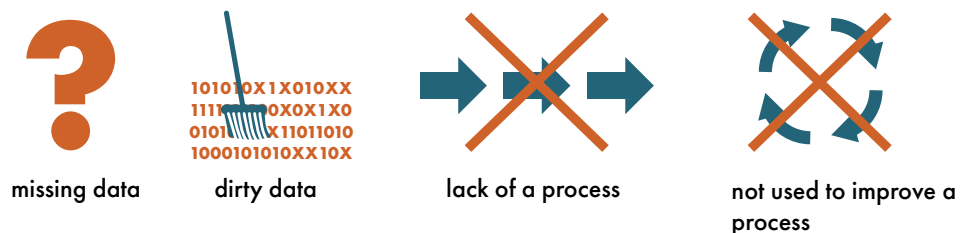
Departments across the City of Austin create and collect vast amounts of data. These data are collected to measure and improve the City's programs, provide services, and to inform community members. To serve these purposes, the data needs to be reliable and easily understandable.

### Previous audits

Over half of our office's audits from 2017 through 2021 identified some type of data reliability issue. These audits covered 20 departments and various Citywide issues. The most common data reliability issues were:

- Data either missing or not collected
- "Dirty data," which includes poor data quality, misspellings, extra characters, duplicates, or outdated data
- Lack of process for collecting or evaluating data
- Departments not using data to improve processes

**Exhibit 1. The most common data reliability issues.**



Source: OCA analysis of prior OCA audits, 2017-2021.

For this audit, we sampled seven high risk audits our office previously conducted to determine if some of the issues we initially found had been addressed. Departments made progress with their data issues in five instances. Of those, only three were fully implemented. In one instance, the department created new data issues through how they addressed the recommendation. About half of the data reliability recommendations we sampled were behind schedule.

**Exhibit 2. Four of seven past audits still have unresolved data reliability issues.**

Audit	Department	Original Data Issue	Were data issues addressed?
APD Response to Mental Health-Related Incidents (2018)	Austin Police	Department does not analyze data or use it to improve processes	Yes
APD Body-Worn Cameras (2019)	Austin Police	Department does not analyze data	Yes
APD's Early Intervention System for Officers (2021)	Austin Police	Collecting wrong data	No
Homelessness Assistance Series: Outcome of City Efforts (2019)	Austin Public Health	Dirty data	Partially
Access to Mental Health Services (2019)	Austin Public Health	Data not collected or missing	No
City Coordination in the Right of Way (2018)	Austin Transportation	Inconsistent process for recording right-of-way inspections	Yes
Audit of the City's Harassment, Discrimination, and Retaliation Investigation Practices (2017)	Human Resources	Software is inadequate for managing cases and producing reports	Partially

Source: OCA analysis of prior OCA audits, 2017-2021.

As a result of the data reliability issues we found in previous audits, we decided to look at the City's data reliability more broadly. We focused this audit on the City's Open Data Portal.

## Open Data Portal

The City's Open Data Portal is a public data sharing website that provides community members access to a wide range of the City's data. The portal's data covers topics across City operations. These include animal care, permitting, health, transportation, public safety, and other topics. Departments upload data, called "assets," to the portal. Some data are manually uploaded; others are uploaded through automated processes. The portal contained over 4,000 assets at the time of this audit, including:

- Datasets, which are the raw data organized in spreadsheets
- Data stories, which often provide a written narration with charts, graphs, or photographs to provide context for data
- Standalone charts and graphs
- Maps

Open Data Portal users can download data from the portal and connect it to other analytical tools. Depending on the type of data and how the data are formatted, users may also use the portal's visualization tools to create graphs and maps. The data are also "machine readable," which allows computer processing of datasets.

Community members requested an open data portal as part of the City's website redesign in 2011 and 2012. The City began a portal pilot program shortly after. The City's 2013 Open Data Directive expanded the scope of the portal and directed departments to initially identify and publish three high-value datasets each. Communications and Technology Management (CTM) is the department responsible for managing the portal.

---

## What We Found

### Summary

The City lacks a strategy for providing data to the public. The Open Data Portal contains over 4,000 assets, which include a mixture of datasets, charts, maps, filtered views, data stories, and other files. However, many assets are specific to past events or initiatives and are outdated. Most assets appear to be of low interest to the community and are not frequently accessed. Some have been accessed only once. We also found that datasets are not consistently formatted or easy to interpret. Datasets are not always formatted so users can use the portal's built-in visualization tools.

We found the data on the portal did not consistently match departments' data sources. The discrepancies between data on the portal and in department sources varied from as few as two missing records to hundreds of thousands of missing records. This means community members and City decision makers who use data from the portal may be getting information that is incomplete, inaccurate, or otherwise different from the data departments may use when making decisions.

# Finding 1

The City lacks a strategy for providing data. Individual departments make data publicly available through the Open Data Portal, but often without a clear purpose or audience for the data. Data is often not presented in a way to make it accessible to a wide audience.

**Exhibit 3. The City meets some of the best practices for open data and some improvements are underway.**



Source: OCA analysis of open data best practices.

The City follows some of the best practices for open data and is in the process of improving on others. However, the City does not have a clearly identified audience for its data and does not have an overall strategy for what data departments should share. The 2011 City Council resolution and 2013 Open Data Directive that led to the Open Data Portal say departments should make as much data as possible available in primary forms but does not provide additional guidance. As a result, the portal contains a mishmash of data, much of which is outdated.

“Proactively disclosing City data is the foundation of Open Government, is consistent with citizens’ right to public information, and promotes engagement with the potential benefit of civic development of applications to improve service delivery through expanded and innovative applications.”

— City of Austin City Council Resolution

## Data on the portal are not overseen by any one department or person

Communications and Technology Management (CTM) is the department responsible for managing the technical aspects of the Open Data Portal. Yet, CTM is not responsible for the content of the portal itself. CTM is also not responsible for a strategy of what data to share or how data should be shared to reach a wider audience. The City’s Strategic Direction 2023 includes indicators departments report on through the portal, but otherwise, department liaisons are responsible for selecting and uploading assets to the portal.<sup>1</sup> We looked at other cities’ data platforms and found many similarities to Austin’s portal. However, unlike most of the other cities we looked at, Austin does not have a “Chief Data Officer” or “Chief Analytics Officer,” whose role is to oversee their city’s data and analytics initiatives.

## The City does not have an inventory of assets

Keeping an inventory of assets is a best practice for open data portals, but Austin does not have one. The portal contains over 4,000 assets. Austin has more assets than any of the other cities we looked at, including over 1,000 more assets than New York City, the next-most prolific city. However, Austin does not have a good way of tracking and managing those assets. This results in pages of outdated assets on the portal.

## Most assets on the portal have a low number of views

Many of the assets on the portal are of low interest to the community and are not frequently accessed. Some of Austin’s assets have only a single view, indicating people are not accessing the data. CTM has a single performance indicator for the portal: “Percentage of datasets published in the Open Data portal that are being utilized frequently.” CTM considers assets accessed over 1,200 times per year higher interest. In 2021, CTM had a target of 15% of assets meeting or surpassing that threshold. Only 12% of assets met or surpassed the threshold. Forty-two percent of assets have fewer than 100 overall views.

<sup>1</sup> Strategic Direction 2023 (SD23) is the City’s 5-year strategic plan that focused on improving quality of life and civic participation.

# 42%

**of assets have fewer than 100 views.**

Source: OCA calculation, June 2023.

Community members “often lack the time, technical literacy and operational capacity to identify highly valuable open data projects and deeply engage in open data initiatives.”

— Ash Center and the City of Cambridge *Amplifying Civic Innovation* report

## We do not know who is accessing data on the portal

Staff lack the ability to see who is accessing the portal. The portal is mostly intended for the public, but City staff also retrieve information from the portal. However, the portal’s analytics do not allow the Open Data Portal team to see if data is accessed from inside or outside the City’s network. The portal also does not show whether data is being accessed by a human or through an automated process. Not knowing who the City is reaching through the portal makes it hard for the City to assess whether it is meeting its goals of transparency and meaningful public participation.

## The portal does not make data accessible to most people

Providing data is not the same as providing information. Open Data principles adopted by the City make providing as much data as possible in a raw format a priority. But to make use of the data, community members must download and analyze a dataset they may not have the tools, understanding, or knowledge to evaluate. Presenting data through charts, graphs, or other visualizations can engage more people. The City’s current approach can serve users with the time, education, and resources to dig into the data, but it neglects less data-savvy community members. The City can make data more accessible through user-friendly visualizations that explain the data, including graphs, maps, and tables, while still providing the underlying data for people who wish to do their own analyses.

Several issues may impede a person’s ability to analyze data from the portal:

- Many of the datasets exceeded Excel’s maximum number of rows (1,048,576), meaning people using Excel—one of the most popular and widely available data tools—cannot analyze the whole dataset.
- Columns were not clearly labeled in several of the datasets.
- Some data with addresses or geographic coordinates were not set up to be mappable using the portal’s mapping visualization.

Austin can also serve the community better by providing training. Austin relies primarily on documentation created by its portal vendor. Most other cities we looked at provide training for residents and city staff.

**Exhibit 4: Austin provides less user training and support than other cities.**

City	USER TRAININGS			USER SUPPORT
	Online Resources	City Created Resources for the Public	Instructor Led Classes	Frequently Asked Questions
Austin, TX	Yes	No	No	No
Dallas, TX	Yes	No	Yes	Yes
Cincinnati, OH	Yes	Yes	No	No
New York City, NY	Yes	Yes	Yes	Yes
Philadelphia, PA	Yes	Yes	Yes	Yes
San Francisco, CA	Yes	Yes	Yes	Yes

Source: OCA analysis of other cities’ open data portals.

**Exhibit 5. Austin lacks an executive position overseeing open data, an inventory of assets, and standards for open data.**

City	Number of assets	Does the city have a...					
		Chief Data/ Analytics Officer?	publicly available asset inventory?	policy stipulating an inventory?	defined and thorough standard for data?	clear goal, mission, or vision?	policy in code, resolution, or executive order for open data?
Austin, TX	4,672	No	No	No	No	Yes	Yes
Dallas, TX	1,081	Yes	No	Yes	No	Yes	Yes
Cincinnati, OH	158	Yes	No	Yes	No	Yes	Yes
New York City, NY	3,582	Yes	Yes	Yes	Yes	Yes	Yes
Philadelphia, PA	384	Yes	Yes	Yes	Yes	Yes	Yes
San Francisco, CA	1,09	Yes	Yes	Yes	Yes	Yes	Yes

Source: OCA analysis of other cities' open data portals.

Metadata are incomplete or not useful

Metadata provide information to users to help them interpret data. Examples are the source of the data, when the data were updated, who to contact about the data, and the frequency of updates to the data.

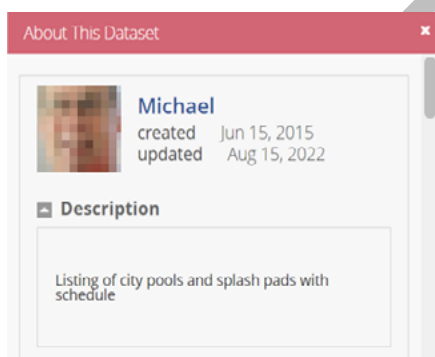
Assets on Austin's Open Data Portal generally adhere to best practices for metadata but have room for improvement. We compared Austin's metadata to best practices and other cities and found that Austin provided a similar level of metadata. All the top assets we reviewed had the metadata considered *always required* by the federal government's open data portal, such as title, description, data owner, and updated date, although those metadata were not always useful. For example, one asset had a data owner given as a first name without additional context.

Other metadata on the portal are sometimes missing or incomplete. The least consistently provided metadata in our analysis were column descriptions and frequency of updates. This hinders people's ability to interpret the data or to see how current the data is.

Other cities did a better job of providing guidebooks or manuals on how users should use their portals, including information on data quality guidelines. CTM has added documentation for the portal since our audit began. They recently updated metadata standards for the portal. This includes requiring certain metadata fields to be filled in before an asset can be uploaded to the portal.

The City's portal also connects directly to datasets hosted on the State of Texas' open data portal. This data is not verified by the City, and the City does not have guidelines for what external data they connect to. According to CTM staff, they are in the process of determining how data hosted elsewhere should be housed on the City's portal.

**Exhibit 6. Who's Michael? Metadata was sometimes included but not useful.**



Source: Screenshot of "About this Dataset" for the Pools Map. Image was pixelated for privacy.



## Finding 2

Data on the Open Data Portal do not always match departments' data. Departments are responsible for the accuracy, timeliness, and usefulness of content, but managing open data is not a primary function of departments.

### Open data do not match departments' data

Departments collect, create, and store data using a variety of software. Those are departments' primary records of information and are separate from assets on the Open Data Portal that are uploaded to be shared with the public. Our office looked at 11 of the most accessed assets on the portal. Nine were datasets, one was a chart, and one was a map.<sup>2</sup> We compared whole datasets when possible but took random samples in three instances because of constraints in accessing the original data sources. We found that data in six of the assets did not match department data. The differences varied from as few as two missing records to hundreds of thousands of missing records. In some instances, we found similar numbers of records, but key data such as dates differed.

#### Exhibit 7. Six out of eleven assets' data do not match the departments' data.

Asset	Do the data match between ODP and originating system?	Types of differences
Crime Reports ( <i>sample</i> )	Yes	None <sup>3</sup>
Mixed Beverage Gross Receipts ( <i>external</i> )	Austin's portal links to the State's portal	None
Top-10-Searches-Chart	Yes	None <sup>4</sup>
Unclaimed Property	Yes	None
AFO eCheckbook (2020-2021 only)	Yes	None
Austin 311 Public Data	No	Missing service requests
Austin Animal Center Outcomes	No	Missing two animals
Issued Construction Permits ( <i>sample</i> )	No	Different permit dates
Food Establishment Inspection Scores ( <i>sample</i> )	No	Scores did not match
Pool Map	No	Pool statuses and hours were outdated
Real-Time Traffic Incident Reports	No	Missing incidents

Source: OCA analysis of Open Data Portal assets.

<sup>2</sup> Both the chart and the map are based on datasets on the portal.

<sup>3</sup> This dataset had a discrepancy in part of their location data, but APD staff independently identified and corrected this issue while we were conducting this audit.

<sup>4</sup> We noted other data reliability issues, which are discussed later in this report.

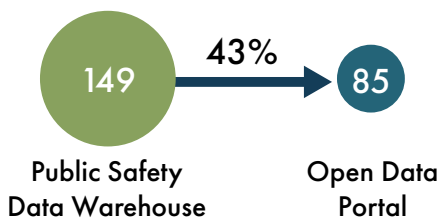


We talked to people in the respective departments when we discovered discrepancies between portal data and the department's data. In a few cases, staff told us data was excluded from the portal for reasons like proprietary information, privacy concerns, or evolving reporting practices. For instance, the Austin 311 Public Data asset does not include certain requests. Even accounting for those exclusions, the data do not match. Moreover, datasets do not always tell users what data may be excluded or modified. Users would not always know that the data on the portal is different from the data in the City's systems of record.

The differences between the data on the portal and the department data we looked at varied considerably. Some differences are minor. For instance, a few Food Establishment Inspection Scores in our sample were off by a couple points: a restaurant received a score of 86 in the department data versus 88 on the Open Data Portal. This points to a reliability problem, but likely has minor effects. Similarly, the Austin Animal Center Outcomes dataset on the portal was missing two animals out of almost 150,000.

Other differences could have larger policy implications. The number of records on the portal for both the Austin 311 Public Data and the Real-Time Traffic Incident Reports are substantially lower than the source data. The 311 dataset appears to be missing hundreds of thousands of mobile, web, and phone service requests, limiting the public's ability to understand the types of requests and how requests are made through 311. The Traffic Incident Reports data on the portal is missing thousands of records. It also shows a different number of crashes, including severe crashes. Someone using the portal data would see 85 fatal crashes between September 2017 and November 2022 when someone using the Public Safety Data Warehouse would see 149. However, Transportation and Public Works staff said they do not rely on either dataset for engineering or project development. Instead, they use a more comprehensive dataset with official state crash reports and publicly share their analyses through a different portal dataset and connected dashboard.<sup>5</sup>

**Exhibit 8. The portal's Real-Time Traffic Incidents Reports shows 43% fewer fatal crashes than in their system of record.**



Source: OCA analysis of Real-Time Traffic Incident Reports and Public Safety Data Warehouse data.

#### There is no oversight or quality control of assets

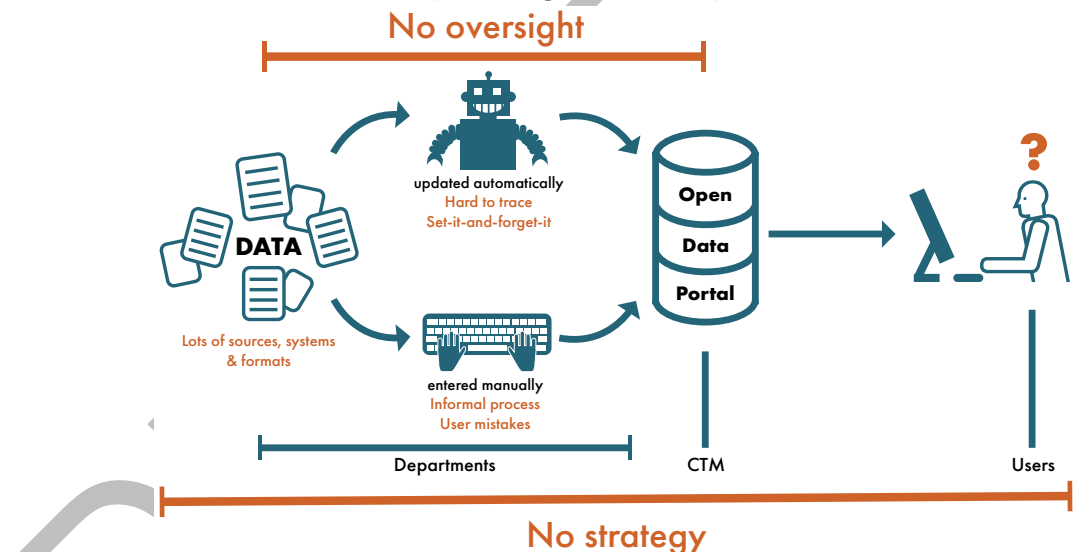
Communications and Public Information Office's (CPIO) Top-10-Searches-Chart reveals a different kind of reliability issue. The chart shows the most visited City of Austin webpages. We found that while the data match, there was no quality control. On multiple dates, we found duplicate search terms displayed in the chart. The search terms were the same aside from a capitalized letter. For instance, "Jobs" and "jobs" were included separately. CPIO staff told us someone from their team manually pulls data from Google Analytics and enters it into the dataset populating the chart. Staff said the data is entered exactly as it is reported by Google Analytics. The team member is responsible for checking the units and dates. However, counting duplicates is misleading because the search term is the same. Combining counts would show the actual popularity of the search term.

<sup>5</sup> This audit compared data on the portal to data used by departments but did not look at the veracity of the data itself. Crash data in particular has some methodological challenges due to differences between jurisdictions, first responders, and reporting systems.

We also found identical data in two different months, suggesting the chart was not updated despite metadata that said it had been.

Both manual and automated assets had missing or incorrect data. Manual updates are an informal process that may rely on staff's institutional knowledge, meaning staff turnover can result in lapses. They are subject to typos or other entry errors and may not be updated regularly. Automated updates also have issues. Staff cannot easily trace data from the department's system of record to the portal. Based on our review, once a transfer is set up, staff are not checking to ensure data displays correctly on the portal or that the data match. Monitoring open data for discrepancies is not something any department we looked at currently does.<sup>6</sup>

**Exhibit 9. Departments are responsible for putting data on the portal. CTM is responsible for managing the portal's operation. No one person or department is responsible for verifying data or an overall strategy for providing data to the public.**



Source: OCA analysis of Open Data Portal assets.

### Managing open data is not a primary function of departments

Department staff were unaware the data on the portal did not match their data. This was true for both manually updated data and automated data. Staff said they did not have the time or the resources to check data to ensure it matched.

Reliable data allows the City to measure its operations and ensure they are effective. It also provides community members a window into City operations and helps provide accountability. Except for the Communications and Public Information Office, none of the sampled departments or CTM had performance indicators addressing data management. Department staff said managing data on the Open Data Portal was not one of their primary job functions.

<sup>6</sup> Financial Services staff said they used to monitor the Open Data Portal to make sure data matched their financial system. Financial Services staff said they have monitoring functions for their system of record.

---

# Recommendations and Management Response

1

The City Manager should articulate a clear goal for open data and data practices more generally. The City Manager should create a Citywide strategy for collecting and sharing data. While the motivation for the Open Data Portal was to make data directly available to the public, a Citywide strategy should:

- Use the performance indicator for how often assets are accessed to determine which assets are higher interest.
- Interpret the most desired data with visualizations and/or data stories, in addition to providing the raw data.
- Remove outdated assets from the Open Data Portal.

**Management Response:** Agree

**Proposed Implementation Plan:**

City of Austin is committed to the open data and transparency principles as established in the City of Austin Open Government Directive (August 2013). While department directors are accountable for their department's data, the City Manager has assigned oversight of Open Data Portal activities to the Communications and Technology Management (CTM) department. CTM will continue implementing open data practices and quality standards across the city. CTM has already developed and published roles and responsibilities for data liaisons and metadata standards for open data. The City of Austin open data portal was revamped earlier this year with additional videos and tutorials on how to use the open data portal. The city has focused on promoting open data and making it available to the public. As the program matured, there has been a conscious effort to make open data more accessible through data visualizations and stories.

- Department directors are accountable for their department data. Department directors will be required to identify department open data liaisons skilled in data management and data analysis whose job responsibilities include publishing and maintaining accurate data for their department.

CTM continues to promote awareness and provides opportunities to city staff through other channels such as 360 Learning sessions. The most recent session was The Art of Storytelling Through Data and Visualizations. CTM has established a training program on open data standards and best practices, including data analysis, extractions, management, and visualizations for department data analysts. The first training session was provided in June 2023.

- CTM will require department open data liaisons to be trained in open data analysis and visualization.
- CTM will sponsor semi-annual training engagements for department data analysts.
- CTM will identify the top three (3) most accessed datasets on the open data portal on an annual basis and will assist departments with data visualizations of the top three datasets if requested.
- Departments will be required to provide stories and visualizations for most-accessed datasets.
- CTM will provide a performance indicator dashboard (refreshed monthly) on open datasets.

---

## Recommendations and Management Response

Removal of outdated assets is an ongoing process. An asset inventory cleanup was completed in February 2023 that removed outdated assets from the data portal. Prior to the cleanup there were 7,274 open data assets. CTM worked with departments and deleted 3,548 unnecessary assets.

- CTM will continue this process semi-annually with the department open data liaisons that will occur once in fall and once in spring.

**Proposed Implementation Date:** July 2024

2

The City Manager should establish centralized oversight over open data, with:

- An inventory of all assets on the Open Data Portal
- Processes in place to ensure data is correct and timely

**Management Response:** Agree

**Proposed Implementation Plan:**

The City Manager has assigned oversight of Open Data Portal activities to the Communications and Technology Management (CTM) department.

As mentioned above, department directors are required to identify department open data liaisons skilled in data management and data analysis who will ensure data is correct and published in a timely fashion. CTM has established a training program on open data standards and best practices, including data analysis, extractions, management, and visualizations for department data analysts.

- CTM will publish an inventory of all assets to the open data portal that will be refreshed monthly.
- Departments must annotate any discrepancies between their source and published dataset or filtering of their source data prior to publication to the open data portal.
- CTM will establish a semi-annual data quality review process.
- Departments will be required to participate in the semi-annual data quality reviews conducted by CTM.

**Proposed Implementation Date:** July 2024

---

## Scope

The audit scope included 1) audits conducted from 2017 to 2022 that noted data reliability issues; and 2) top accessed Open Data Portal datasets, including data from originating department and on the CTM-managed Open Data Portal.

---

## Methodology

To complete this audit, we performed the following steps:

- Reviewed our office's previous audits dating back to 2017 and categorized any data issues we observed in our audit reports
- Conducted follow up tests for a subset of previous audits with noted data reliability issues
- Selected and analyzed a sample of frequently accessed assets on the Open Data Portal and compared data to departments' source data
- Reviewed metadata from the sample of frequently accessed assets
- Interviewed staff from relevant departments
- Reviewed documentation provided by departments for follow-up on prior audits
- Compared Austin's Open Data Portal to other cities' open data portals

---

## Audit Standards

We conducted this performance audit 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.

The Office of the City Auditor was created by the Austin City Charter as an independent office reporting to City Council to help establish accountability and improve City services. We conduct performance audits to review aspects of a City service or program and provide recommendations for improvement.

#### **Audit Team**

Keith Salas, Audit Manager  
Francis Reilly, Auditor-in-Charge  
Elena Purcell  
Mariel Dempster  
Kendall Byers

#### **City Auditor**

Corrie Stokes

#### **Deputy City Auditor**

Jason Hadavi

#### **Office of the City Auditor**

**phone:** (512) 974-2805

**email:** [AustinAuditor@austintexas.gov](mailto:AustinAuditor@austintexas.gov)

**website:** <http://www.austintexas.gov/auditor>



AustinAuditor



@AustinAuditor

Copies of our audit reports are available at  
<http://www.austintexas.gov/page/audit-reports>

*Alternate formats available upon request*