



Texas Open Data Portal PLAYBOOK

Version 1.0

June 2017



Table of Contents

Letter from the Statewide Data Coordinator	2
1. Open Data Overview	3
1.1 What is Open Data?	3
1.2 About the Texas Open Data Portal	4
2. Roles and Responsibilities	5
2.1 DIR Open Data Team	5
2.2 Agencies	5
3. What is an Open Data Inventory?	8
3.1 Guiding Questions	10
3.2 Manual Open Data Publishing	11
3.3 Automated Open Data Publishing	12
3.4 Publishing Process	14
3.5 Data Visualization	14
4. Additional Resources	17
4.1 Portal links	17
4.2 Glossary of Terms	18
4.3 Supporting Legislation	19
4.4 Contact Information	20

Letter from the Statewide Data Coordinator

Open Data is a foundational component of a trusted and reliable government. Constituents invest in their public entities and leadership by voting and have a right to the government's public information. Transparency is the government's obligation to share information with citizens, and lies at the core of how constituents hold their public officials accountable.

A key objective of the Department of Information Resources (DIR) Statewide Data Program is to facilitate and encourage Open Data sharing, through the Texas Enterprise Information Management (TEIM) community group and to leverage the Texas Open Data Portal, at <https://data.texas.gov/>, as the one-stop website to access Texas Open Data. In 2015, the 84th Texas Legislature passed Senate Bill 1844, creating the Interagency Data Transparency Commission (IDTC). The IDTC was formed to review and study the data reporting practices, including Open Data, of Texas state agencies.

During the IDTC's work, the Commission administered a data survey to the executive directors of Texas state agencies. The results of the survey helped the IDTC gain valuable insights regarding the challenges and barriers associated with interagency and Open Data sharing, as well as create a list of recommendations for continued improvement in several areas. One of the recommendations was to establish a formal training and education program for Open Data practices. DIR has initiated several educational and training efforts, including this playbook.

As Statewide Data Coordinator for Texas, I am pleased to present the Texas Open Data Portal Playbook as a resource for current and future participating agencies of the Texas Open Data Portal. This document will serve as a guide for Texas public entities in establishing an Open Data governance framework, creating an Open Data inventory, and publishing Open Data in an efficient and standardized manner. The Texas Open Data Portal Playbook will exist as a living document and will be modified as needed to reflect changes in Open Data best practices and policies.

Sincerely,

Ed Kelly

Statewide Data Coordinator
Office of the Statewide Data Program
Department of Information Resources

1. Open Data Overview

1.1 What is Open Data?

Open Data refers to information that can be freely used, re-used, and redistributed by anyone. This data should be available at no direct cost and is accessible to a wide range of users through Internet downloads.

Open Data is made accessible to the public in a format that is easily read by computers and can be filtered, sorted, manipulated, and combined with other data. Common file types of machine-readable data include, but are not limited to:

Machine-readable data is information in a format that can be processed by a computer.

- Comma Separated Values (.csv)
- Microsoft Excel (.xls/.xlsx)
- Tab Separated Values (.tsv)
- Geospatial data (.kml, .kmz, .json, and ESRI Shapefile)

The Open Government Initiative of 2013 made open and machine-readable data the new default for public information. Government entities are beginning to move to a centric model in which more public datasets are published in machine-readable formats on a central platform. This is an important transition as public entities at all levels of government hold vast amounts of data.

Excluded Data

*Personally Identifiable Information (PII) and Protected Health Information (PHI) are protected by law and **are not** Open Data.*

In defining Open Data, it is important to note what types of information are excluded from this term. Health and personal information protected by law, such as Personally Identifiable Information (PII) and Protected Health Information (PHI) as mandated by the Privacy Act of 1974, are common examples of information that are not Open Data.

Federal law strictly prohibits the sharing of this information with the public and regulates how this data may be shared between government agencies.

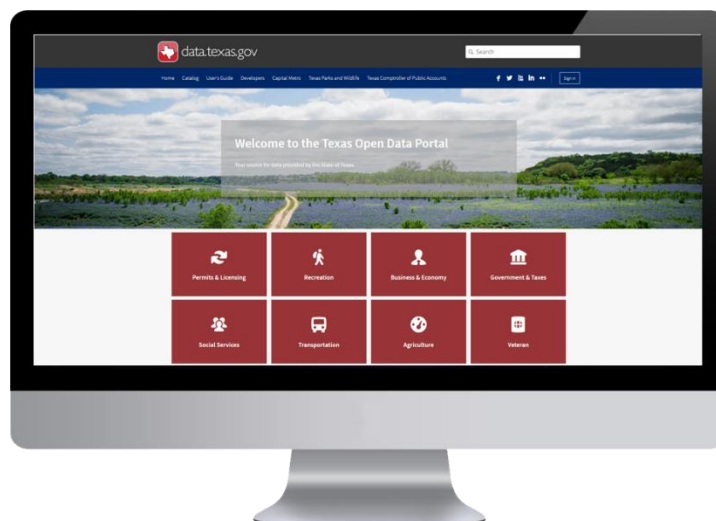
Common examples of PII/PHI data include:

- Names and initials in any combination
- Identification number (i.e. Social Security Number)
- Birthdate
- Age
- Gender
- Home address
- Telephone number
- Personal e-mail address
- Drivers' license number
- Information on medical or health conditions
- Financial information (credit card/PCI, billing info, account info)
- Health information
- Marital status
- Nationality
- Physical characteristics
- Racial or ethnic origin
- Religious, philosophical, or political beliefs
- Biometric data
- Household information
- Consumer purchase or billing history
- Unique device identifiers (IP/ MAC addresses)
- Location (e.g., GPS) info (including that provided by mobile devices)

It is important to note that this is not an exhaustive list. The U.S. Department of Justice provides an overview of the Privacy Act of 1974 as well as additional resources on handling protected personal information at <https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition>.

1.2 About the Texas Open Data Portal

The Texas Open Data Portal, located at <https://data.texas.gov/>, is the official website to access State of Texas Open Data. The portal is built on Socrata's Open Data Portal Platform, a software-as-a-solution (SaaS) developed specifically for government and recognized by Gartner as a predominant Open Data platform.



2. Roles and Responsibilities

2.1 DIR Open Data Team

The Department of Information Resources (DIR) Open Data Team provides leadership and guidance on best practices for participating agencies as well as potential new Open Data publishers. The Open Data Team is responsible for the day-to-day operations management of the Texas Open Data Portal, including:

- Granting access and publishing rights to agencies,
- Managing the dataset inventory,
- Providing guidelines on Open Data governance and publishing, and
- Serving as a liaison between agency publishers and the Open Data cloud provider.

Open Data Portal Advisory Council

To further facilitate collaboration and best practice sharing among Portal publishers, DIR established the **Open Data Portal Advisory Council (ODPAC) in 2017**. The group is comprised of authorized state agency publishers and/or representatives. The ODPAC meets quarterly to discuss how to make improvements to the Texas Open Data Portal, exchange real life success stories related to Open Data sharing, and work with the Statewide Data Program on Open Data policies and related issues. ODPAC meeting minutes, decisions, and updates are shared with the Texas Enterprise Information Management (TEIM) data community group.

2.2 Agencies

Establishing an Open Data Governance Framework

A strong data governance framework is key to maintaining the availability, integrity, and security of an organization's information.

Data governance may be related to an agency's mission, can determine what type of information is available to the public, and how sensitive information is handled.

It should be noted that an Open Data Governance Framework is part of a larger enterprise-wide data governance framework, which pertains to all data types that an organization manages. The Open Data Governance Framework follows similar methods and processes as the larger-scale governance model. The necessary details of your agency's Open Data Governance Framework will depend on the size, volume, and constituent interest.

To establish this Framework, your agency should first develop a structured charter or policy guide that identifies:

- The agency's goals and vision for Open Data
- Roles and responsibilities of Open Data stakeholders and publishers
- Guidelines for classifying Open Data to establish an Open Data inventory
- An approval/vetting process for publishing Open Data to the portal

However, the formality of the charter or policies will vary by agency.

Open Data Stewards

Once an eligible agency has completed the onboarding process (*see Section 3.2*), it must nominate its **Open Data Stewards**. Each steward is responsible for managing the publication of Open Data for a specific program area within the agency. Depending on the size and data assets of the agency, the number of stewards can range from one to more than 10 employees.

Based on DIR's security recommendations, each steward with publishing rights should have a unique Portal login ID linked to his or her government entity-assigned email address. It is not recommended that publishers share Portal IDs or passwords, as this increases the risk of security breaches.

Each Open Data Steward should have a unique Portal ID. Sharing accounts among publishers is discouraged due to increased security risks.

Open Data Stewards serve as key contributors in the following activities, which help improve the Open Data publishing process:

- Identifying Open Data to publish
- Ensuring data accuracy and quality
- Determining the frequency and method of updates
- Reporting and analyzing Open Data consumption

Agency leadership from both business and IT units should collaborate to determine who is best suited to serve as Open Data Stewards, as well as the total number of stewards needed. Below is a list of suggested job titles that can be used to help guide the selection process:

- Database Administrator
- Computer Operations Technician
- Systems Analyst
- Business analyst
- Programmer
- Computer Operations Specialist
- Web Administrator
- Planner
- Research Specialist
- Statistician/Economist
- Program Specialist/Manager
- Information Specialist

Depending on the technical complexity level of Open Data publishing that an agency chooses to pursue, as well as its needs and demands, the skill level needed of each Open Data Steward may vary.

It is critical that agencies understand their roles and responsibilities in publishing to the Open Data Portal. As *Section 3.2* outlines, agencies are solely responsible for the content, accuracy, and maintenance of their Portal datasets, and therefore bear all responsibility in the case of improper release of sensitive data that may have legal ramifications. Therefore, it is key for agencies (and especially publishers) to fully understand the characteristics of sensitive data that is protected by law. These data types are outlined in *Section 1-Open Data Overview*, and additional information about confidential data can be found in *Section 5-Additional Resources*.

3. What is an Open Data Inventory?

As described in *Section 2.2*, each agency participating in the Texas Open Data Portal should establish its own Open Data Governance Framework that is part of a greater enterprise-wide data governance practice. As a core part of this framework, an organization must determine what data is public or “open.” Competing priorities, limited resources, and the often vast amounts of data available also require agencies to prioritize the publication of certain data that will hold more value for constituents.

An Open Data Inventory is exactly that: an itemized list (typically stored electronically as a dataset itself) of available public datasets that the agency has published or may publish in the future.

Other than giving basic context of each dataset, the inventory should also provide information about:

- File type
- Size
- Quality
- Dataset owner
- Update frequency
- Data security/privacy constraints

Metadata

Your agency’s datasets should also include applicable metadata (see *Table 3.1*). **Metadata**, or ‘data about the data’, is a set of data that provides contextual information about each dataset. Metadata plays a vital role in the indexing of each dataset and ensures that all personnel that handle or publish the data follow protocol.

Metadata is a set of data that provides information about each dataset. It makes finding and working with instances of data easier.

The table below shows what kind of metadata should be captured by the inventory but is not exhaustive. Although the Portal currently only requires the fields of Title, Category, and Agency, it is beneficial for each agency to collect more detailed information about each dataset.

Table 3.1: Examples of Metadata for an Open Data Inventory

Dataset Metadata	Common Responses
Title	Dataset title
Category	<ul style="list-style-type: none"> • Agriculture • Business & Economy • Education/Energy • Environment • Government & Taxes • Permits & Licensing • Public Reports & Maps • Public Safety • Recreation • Social Services • Veteran
Agency	Agency name
Description	Varies by dataset, but should include a general description easily understood by the public.
Department	Department name
File Type	.csv, .xls, .xlsx, .tsv
File Size	5 MB
Priority	High, medium, low
Data Owner	Ellen Dataowner
Data Contact	E.dataowner@dir.texas.gov
Quality	Excellent, Good, Fair, Poor
Privacy/PII	Flag only
Security	Public, Sensitive, Confidential
Update Frequency	Daily, Weekly, Monthly, Quarterly, Yearly

3.1 Guiding Questions

As your agency begins to build its own Open Data Inventory, agency personnel must decide which datasets offer the most value to the public and any consumers of the data. The following guiding questions will help your agency identify and prioritize Open Data to create an Open Data Inventory for publication:

- What information is frequently requested by the public, via the Public Information Request (PIR) process?
- What data does your agency or program areas use for key performance indicators (either State Auditor's Office or internal measurements)?
- What data populates your monthly, quarterly, or annual agency reporting requirements?
- What information do you report to local, state, or federal agencies?
- What information do you share with other agencies?
- What information do you share with external stakeholders?
- What kinds of Open Data are your peer agencies across the country publishing?

In addition to the questions above, other factors that must be considered include:

- The quality of the data
- How often new data will be available (update frequency)
- File format (machine-readable datasets are preferred)
- The ability to automate publication after the initial dataset is made public (see *Section 4.2*)

TESTIMONIAL

"The Texas Open Data Portal's proactive approach to delivering data to stakeholders and the public allows them real-time access when they need it most. Fast and efficient access fuels the support and understanding of our mission — feeding the hungry and promoting healthy lifestyles. In addition, this self-service option reduces the amount of staff time used to pull data for Public Information Requests, thereby affording more time for staff to ensure compliance and accountability in the administration of 12 federal nutrition programs in Texas. The Texas Open Data Portal has the potential to become a vital tool in helping TDA Food and Nutrition meet its primary objectives of providing customer service excellence and achieving high professional standards."

Angela Olige, Assistant Commissioner, Texas Department of Agriculture

3.2 Manual Open Data Publishing

Manual publishing is the simplest way to upload a dataset to the Open Data Portal. Using this method, a publisher can upload an individual file (.csv, .xls, .xlsx, .tsv) or create a dataset from scratch. Publishers can also make necessary edits before publishing. Once marked as “public” each dataset can be manually or automatically updated once new data becomes available.

Private vs. Public Data

The Texas Open Data Portal has another useful function for data publishers – the ability to mark a dataset as “private” before it is marked “public”. When a publisher uploads a new dataset, it remains in private mode until the “make dataset public” option is selected. While in private mode, only data publishers and Portal administrators can view or make changes to the dataset. This enables several Open Data Stewards within the same agency to review a dataset before publication, ensuring that data quality, accuracy, privacy and/or security concerns are properly vetted.

Data publishers can also share editing permissions with colleagues within the same agency, to allow for a vetting process of datasets before they are made public. Figure 4.1 explains the types of users, with permissions indicated with an “X”:

Figure 4.1: Open Data Portal User Privileges

		Administrator	Publisher	Editor	Viewer	Public
Dashboards	Create Dashboards	X				
	Edit, Delete Dashboards	X	X			
	View All Dashboards	X	X	X	X	
	View Public Dashboards	X	X	X	X	X
Goals	Edit, Delete All Goals	X	X			
	Create, Edit, Delete Goals	X	X			
	View All Goals	X	X	X	X	
	View Public Goals	X	X	X	X	X
Reports	Edit, Delete All Reports	X	X			
	Create, Edit, Delete Reports	X	X	X		
Data	Edit, Delete All Data	X	X			
	Create, Edit, Delete Data	X	X	X		
	View All Data	X	X		X	
	View Public Data	X	X	X	X	X

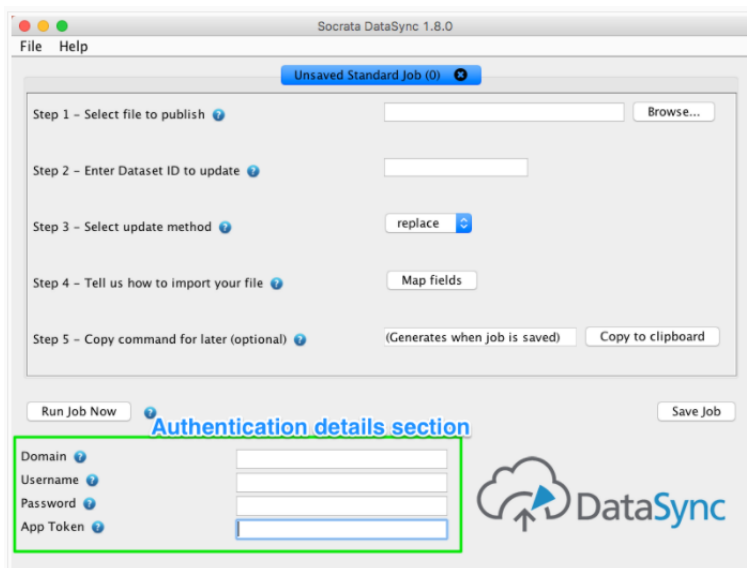
3.3 Automated Open Data Publishing

DataSync

The **Socrata DataSync** tool is a powerful and free publishing tool that lets users schedule and automate data updates and upload large data files on the Texas Open Data Portal. The program requires your computer to be operating on Java v. 7 or higher, and can be downloaded for free from <https://github.com/socrata/datasync/releases>.

When you have downloaded DataSync and open the program, the window will look similar to what is displayed in Figure 4.1 below.

Figure 4.2: DataSync Automated Publishing



To begin the automated publishing process, follow these steps:

Step 1: Select the file you would like to publish from your computer (.csv/.tsv files only).

Step 2: Enter Dataset ID to update (the 8 character code separated by a dash at the end of your dataset's URL).

Step 3: Select update method (option to replace, upsert, append, or delete). See <https://support.socrata.com/hc/en-us/articles/210850597-Getting-Started-with-DataSync#datasync-prefs> for more detailed explanation.

Step 4: Indicate how to import your file (to ensure map field are correctly coded).

Step 5: Copy command for later use (optional, use to copy and save a command to run from your command line or to schedule future jobs).

To complete the Authentication Details Section (highlighted in the image above), enter the following information:

Domain: <https://data.texas.gov>

Username/Password: Your Socrata Open Data Portal login credentials.

App Token: A user-specific code to use DataSync, see <https://support.socrata.com/hc/en-us/articles/210138558> for instructions on generating.

Socrata Open Data API

The Texas Open Data Portal also has tools that allow agencies to keep their own website data synchronized with datasets on the Portal. Each uploaded dataset comes with a built-in system called the **Socrata Open Data API** that allows websites and other types of applications to communicate directly with the Socrata platform via code.

API (Application Programming Interface) is a predefined way to allow one computer program – such as a website – to communicate with another (like the Socrata platform).

This process may take a little extra time for a developer to initially set up, but the benefit is that your website or application will automatically refresh as the Open Data Portal datasets are updated.

Here are a few situations in which it might be worth considering using the API:

- You update data on the ODP and then update the same data on your website as a separate process.
- A dataset on your website must be updated at regular intervals.
- You would like to add more context to a dataset or present a dataset in a manner that is not available on the Socrata platform.
- You have limited time and/or resources to update your Open Data on a consistent, manual basis.

Visit the Socrata Developer Portal at <https://dev.socrata.com> for a better understanding of the API's capabilities, along with technical guidance on how to get started.

TESTIMONIAL

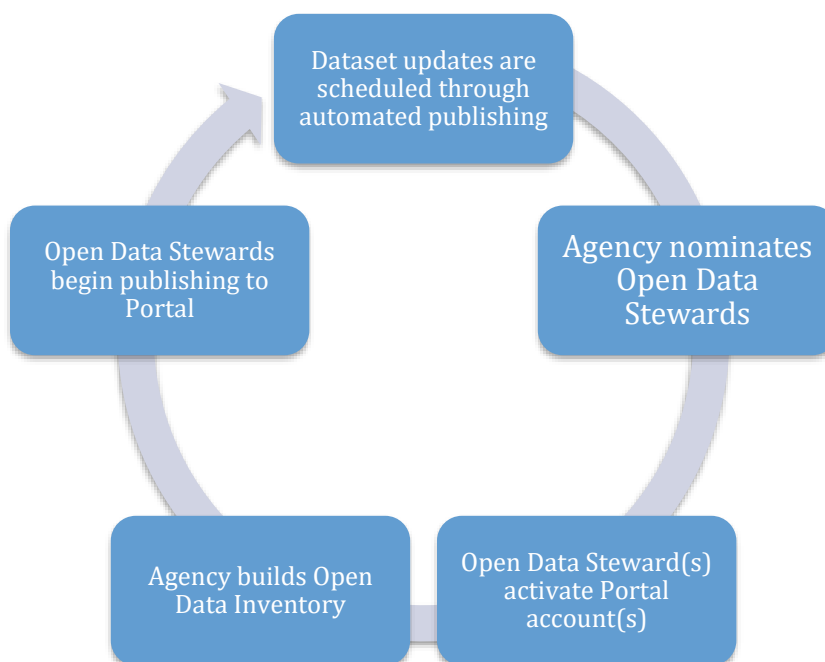
"The Texas Open Data Portal played a key role in modernizing our website. Using Socrata's tools, we were able to increase our efficiency by automating the transfer of data to our website. The portal also boosted our Open Data presence and created new opportunities for collaboration with other state agencies."

Robert Wood, Associate Deputy Comptroller, Operations and Support

3.4 Publishing Process

Whether you are using the manual or automated method to publish datasets to the Texas Open Data Portal, the publishing process should generally follow the cycle illustrated in Figure 4.2.

Figure 4.2: The Open Data Publishing Process



3.5 Data Visualization

In addition to the data publishing methods described above, the Texas Open Data Portal also provides graphic data visualization capabilities for publishers and viewers alike. This section provides general guidance in using these functions, but more detailed information, tutorials, and videos can be found at <https://support.socrata.com/hc/en-us/sections/200469728-Data-Visualizations>.

Charts, Graphs, and Maps

The basic graphing, charting, and mapping capabilities can be leveraged by any type of user (viewer, publisher, administrator). If a user has created a Socrata account and is logged in, he or she can create and save filtered views, maps, or charts. Publishers can also use these features to graphically represent datasets and publish them as “official views” so that they appear in the data catalog.

These features include:

Filtering and Sorting:

- Filtering and sorting
- Conditional formatting
- Roll-ups and drill-downs

Charts, Graphs and Maps:

- Calendar mapping (column data must be formatted as date)
- Mapping (column with address, zip code, or city must be formatted as location)
- Charts, depending on data type (column, bar, pie, donut, line, area, timeline, bubble and tree maps)

Some of these functions are displayed in the graphics below.

Figure 4.3: Clustered Bar Graph

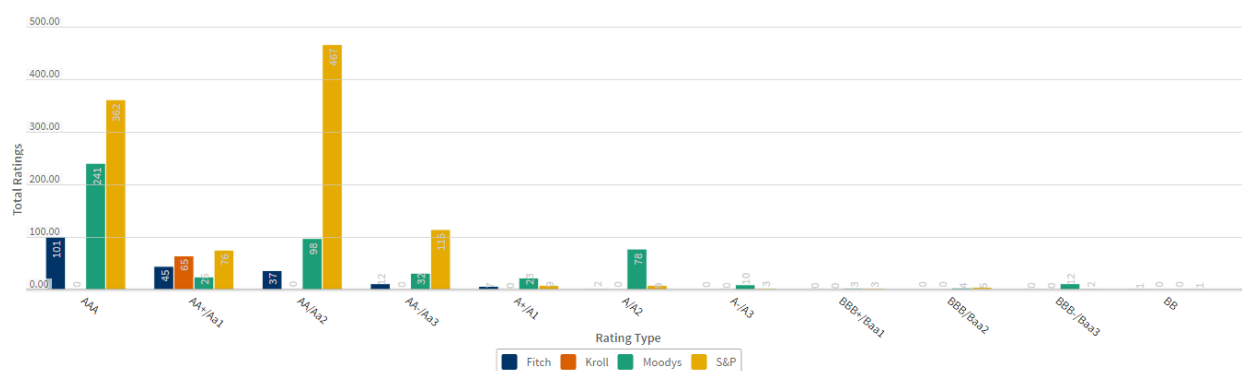
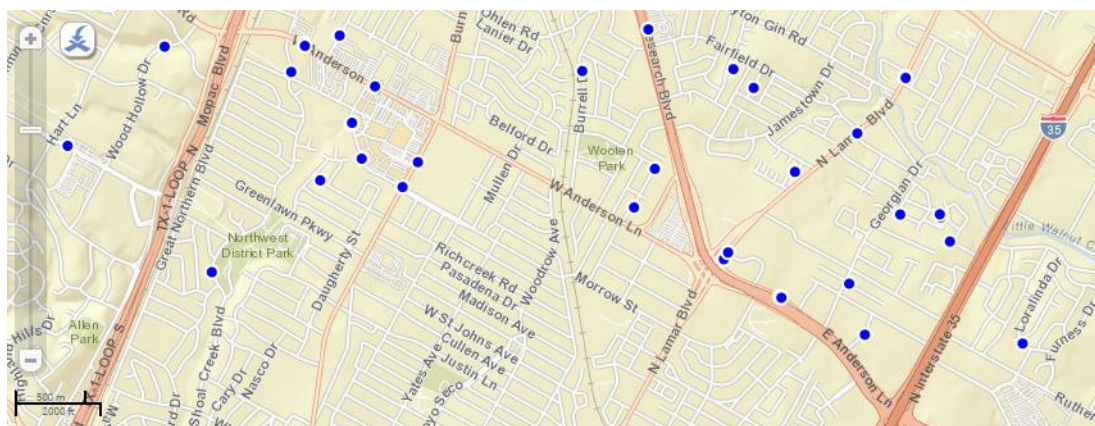


Figure 4.4: Point Map by Location



Data Lens

The **Socrata Data Lens** tool allows Texas Open Data Portal publishers to create data visualizations that the public can easily engage and interact with, by customizing how data is presented in the catalog. Datasets with geographic information can be mapped both by point location and by aggregate comparisons by county, Councils of Government, Senate, and House Districts.

Other functions include search bars by column and bar graphs. Figures 4.5 and 4.6 show examples of Data Lens visualizations. These engaging visuals can also be exported to another website via API or downloaded as a .png file.

Figure 4.5: Data Lens Heat Map

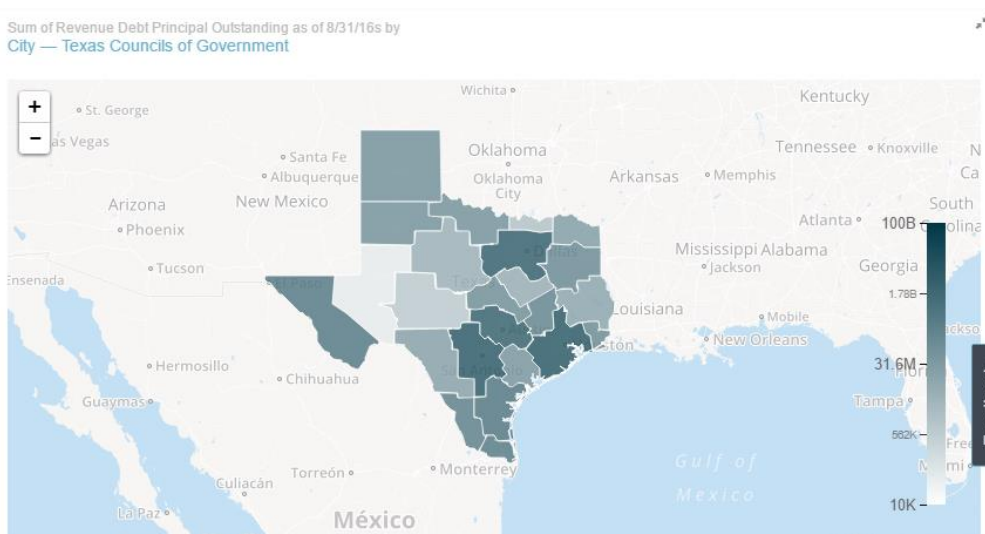
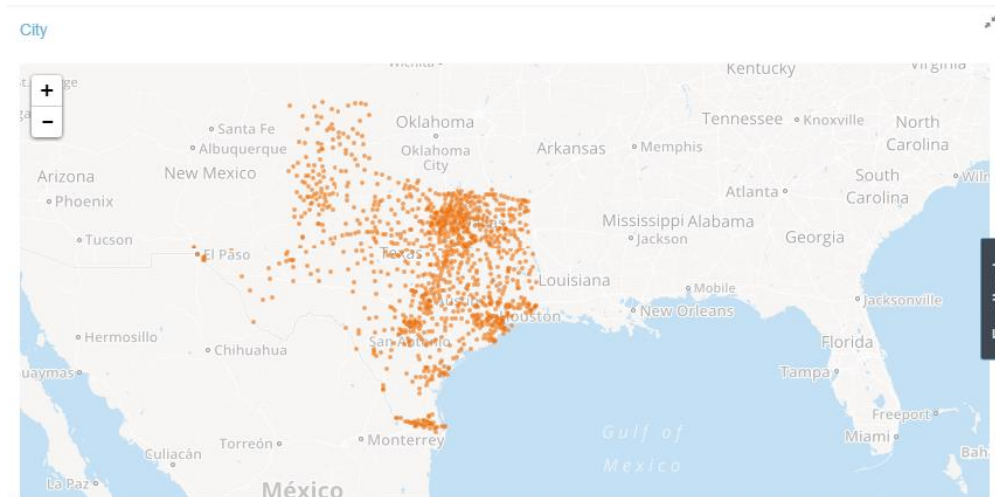


Figure 4.6: Data Lens Point Map



4. Additional Resources

4.1 Portal links

- Socrata General Support: <https://support.socrata.com/hc/en-us>
- Socrata Developer Portal: <https://dev.socrata.com>
- Federal Open Data Portal: <https://www.data.gov/>
- City of Austin Open Data Portal: <https://data.austintexas.gov/>
- Privacy Act of 1974: <https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition>

4.2 Glossary of Terms

- API – Application Program Interface. API is programming code that allows for two distinct applications to communicate with each other in an automated way.
- Data – information that has been translated into binary components so that computers can move, store and process it.
- DataSync –the Socrata proprietary tool that has been developed to automate the data publishing process for Socrata platforms.
- Open Data Steward – the person or persons responsible for management and oversight of an agency’s Open Data. They provide a vital role is ensuring the quality, availability, and consistency of the data.
- Machine-readable – data that is in a standard and format that can be easily recognized and processed by computers.
- Metadata – ‘data about the data’. Metadata provides the contextual understanding of each dataset and creates an index for all datasets so they can be easily found.
- Open Data –data that refers to information that can be freely used, re-used, and redistributed by anyone.
- Open Data Governance Framework - part of the larger data governance framework of people, process, and technology to manage an agency’s information. The framework for Open Data is the key to maintaining the availability, integrity, and security of an organization’s information. Open Data Inventory is an itemized list (typically stored electronically as a dataset itself) of available public datasets that the agency has published or may publish in the future.
- Texas Public Information Act – the State’s law that entitles each citizen, unless otherwise expressly provided, to have access to information about the affairs of government and the official acts of public officials and employees.

4.3 Supporting Legislation

Texas Public Information Act – Title 5, Open Government, Ethics, Chapter 552 - Public Information (September 1, 1993)

Set forth the policy of the state that each person is entitled, unless otherwise expressly provided by law, at all times to complete information about the affairs of government and the official acts of public officials and employees.

<http://www.statutes.legis.state.tx.us/Docs/GV/htm/GV.552.htm>

Open Data – SB701 82(R) (September 1, 2011) Required state agencies to post high-value data sets on a generally accessible agency website. High-value data sets include, but are not limited to, data that is critical to the financial and programmatic function of state agencies.

<http://www.capitol.state.tx.us/tlodocs/82R/billtext/pdf/SB00701F.pdf#navpanes=0>

Open Data – SB279 83(R) (September 1, 2013) Required state agencies to send the Department of Information Resources (DIR) a description and link to their high value data sets. <http://www.capitol.state.tx.us/tlodocs/83R/billtext/pdf/SB00279F.pdf#navpanes=0>

Statewide Data Coordinator – HB1912 84(R) (September 1, 2015) Created position to collaboratively develop data policies, standards, and best practices and to improve data governance and integrity statewide.

<http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/HB01912F.pdf#navpanes=0>

Interagency Data Transparency Commission – SB 1844 84 (R) (September 1, 2015)

Created the commission to study and review the current public data structure, classification, sharing, and reporting protocols for state agencies.

<http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/SB01844F.pdf#navpanes=0>

Legacy Modernization/Data Reporting - HB1890 84(R) (September 1, 2015) Relating to the development and implementation of a statewide strategy for legacy system modernization. Also, implementation of a pilot shared data reporting and business analytics service.

<http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/HB01890F.pdf#navpanes=0>

4.4 Contact Information

Please contact the DIR Open Data Team below with any questions regarding the Texas Open Data Portal:

Ed Kelly, Statewide Data Coordinator

ed.kelly@dir.texas.gov

(512) 463-1811

Ingrid Olsson, Digital Government Business Analyst

ingrid.olsson@dir.texas.gov

(512) 475-4526