Texas Data Portals

Resource Guide | January 2022

Texas Data Portals Resource Guide

Office of the Chief Data Officer
Texas Department of Information Resources

Acknowledgements

Table of Contents

Welcome from the Texas Data Management Program

Open data is a critical component of a trusted and reliable government. Constituents look to their government to provide transparency of business and operational practices. This transparency allows constituents to hold their public officials accountable to serve in the best interests of the public good.

In 2018, the 86th (R) Texas Legislature introduced Senate Bill (SB) 819 which focused language on the continued importance of open data and the Texas Open Data Portal. On September 1st, 2019, the bill became law. SB 819 states that the Department of Information Resources (DIR) "shall establish a central repository of publicly accessible electronic data as the official open data Internet website for this state" and that the repository shall be designated as the Texas Open Data Portal (ODP). Most recently, the 87th Legislative Session introduced SB 475 directing each Texas state agency with more than 150 full-time employees (FTEs) to post a minimum of three (3) high value datasets to the Open Data Portal. Signed into law in June 2021, SB 475 further reinforces the significance of open data.

The Office of the Chief Data Officer (OCDO) recognizes the importance and value of the ODP and provides state agencies, political subdivisions, and institutions of higher education shared access to the repository to post publicly accessible open data. In leveraging the ODP, and working together, we can ensure that the business of our government is transparent and performing up to the expectations of those whom we serve.

Sincerely,

Ed Kelly

Chief Data Officer, Texas Department of Information Resources

---

One of the five key service areas of the Texas Data Management Program (TDMP) is guidance, training, and education in data literacy. To meet this need, our program has published the Texas Data Management Framework guide and regularly offers workshops and conferences related to data management best practices. The Data Portals Resource Guide is a part of the ongoing data literacy activities of the TDMP and provides training and education about the Texas Open Data Portal.

While not exhaustive, this document serves as a guide for Texas public entities in establishing an open data governance framework, creating an open data inventory, publishing open data in an efficient and standardized manner, and leveraging data sharing tools. This resource guide exists as a living document with links for additional guidance. It will be modified as needed to reflect changes in data policies and best practices.

Sincerely,

Stacey Lewis

Open Data Portal Administrator, Texas Department of Information Resources

1. Common Terms

Below are commonly used terms you will find in this guide. Additional terms may be found in the Glossary.

- **Agency:** For the purposes of this guide, the term "agency" is generally used to indicate a state agency or a state institution of higher education.
- **CDP:** Closed Data Portal, a private sharing environment administered by the agency
- **OCDO:** Office of the Chief Data Officer at the Department of Information Resources
- **ODP:** Texas Open Data Portal, administered by the OCDO Open Data Portal Team
- **ODPUG:** Open Data Portal User Group

2. Open Data Overview

2.1 What is Open Data?

Open data refers to information that can be freely used, reused, and redistributed by anyone. Such data should be available at no direct cost and accessible to a wide range of users, typically through internet downloads. Open data is made accessible to the public in a format that is easily consumed to be read, 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

- .csv (Comma Separated Values)
- .xml (Extensible Markup Language)
- .tsv (Tab Separated Values)
- geospatial data (.kml, .kmz, .json, ESRI Shapefile)

The federal Open Government Initiative of 2013 made open and machine-readable data the 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 – an important transition as government entities of all levels hold vast amounts of data.

Effective September 1, 2019, the Texas Legislature 86th (R), through Senate Bill (SB) 819, ushered in important advances for open data in the State of Texas. SB 819 amended the Texas Government Code (TGC), Chapter 2054, to require the establishment of a central open data repository, later named the Texas Open Data Portal (ODP). Under Section 2054.070, Texas Government Code, state agencies and political subdivisions are granted shared access to the repository, which offers an easy way to publish publicly accessible information.

The 87th (R) Legislative Session further highlighted the importance of open data by passing Senate Bill 475, which directs each Texas state agency with more than 150 full-time employees (FTE) to post a minimum of three (3) high value datasets to the Open Data Portal. SB 475 was signed into law on June 14, 2021.

Excluded 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.

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

Other examples include law enforcement data in compliance with the Criminal Justice Information Services (CJIS), education data in compliance with the Family Educational Rights and Privacy Act (FERPA), or any other data that agencies and government entities have deemed private or sensitive in nature. Also excluded from open data is data that state agencies have submitted to the Office of the Attorney General (OAG) and received approval not to disclose due to its sensitive nature.

Federal law strictly prohibits the sharing of this information with the public and regulates how this data may be shared between government agencies. For information on sharing sensitive, confidential, or protected data within the Closed Data Portal (CDP), please see *Section 10*.

Common examples of sensitive, confidential, or protected PII/PHI data include:

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

It is important to note that this is not an exhaustive list. For more information as well as additional resources on handling protected personal information, visit the U.S. Department of Justice's page about the **Privacy Act of 1974**.

2.2 The Value of Open Data

Open data is a public resource that offers both governmental and private entities economic, performance, and social value. The economic value of open data is that it allows private business to better understand potential markets and integrate government data into new, more innovative products and services. Open data also adds performance value by increasing government transparency and accountability to its citizens. Greater openness of government improves the quality and efficiency of government services by allowing citizens and policy makers to make more informed data-driven decisions.

Providing greater visibility into how government manages resources in turn adds social value to open data by building stronger partnerships between citizens and their government. Having free and open access to information is the hallmark of a healthy democracy and encourages greater participation by citizens who feel empowered to give feedback and suggest improvements on how their government is working. To summarize, the value of open data is that it offers:
- economic value by encouraging innovation in products and services,
- performance value by improving quality and efficiency of government programs, and
- social value by increasing citizen participation in government.

Reduction of Public Information Requests

Another important value of open data to Texas governmental bodies is its ability to fulfill citizen public information requests. With the objective of creating a more open government that serves the information needs of an informed citizenry, the Texas Legislature (63R), 1973, passed the Public Information Act (PIA). The PIA gives citizens the right to inspect or make copies of open government records in the form of a public information request (PIR).

Under the PIA, governmental bodies are generally required to release information to citizens unless the information is excepted from disclosure. Governmental bodies must respond to a PIR promptly, generally within 10 business days. Responding to a PIR, however, is both time and resource intensive. While citizens may be charged a fee for fulfilling a PIR, governmental bodies pay a higher cost of lost opportunity time to focus on other core business activities.

> **Texas Government Code, Chapter 552**
> The people, in delegating authority, do not give their public servants the right to decide what is good for the people to know and what is not good for them to know. The people insist on remaining informed so that they may retain control over the instruments they have created.

In 2017, the 85th Texas Legislature passed SB 79, to allow governmental agencies to redirect a PIR to information that is available on sites on the Internet, such as the ODP. Redirecting constituents to the ODP is a great economic and performance benefit to governmental bodies as it reduces the number of PIRs, thus saving time and resources.

For more information, visit the Texas Office of the Attorney General's page about **public information requests.**

Self-Service Model

Publishing open data on the ODP is not only a benefit to governmental bodies but is also of value to constituents. While citizens retain the right to request information directly from a governmental entity, filling out a formal public information request is no longer the only way to access publicly available information. Citizens can now take advantage of open data published on the ODP to self-serve their information needs. In the same way that redirecting a PIR to open sources of information saves governmental bodies time and resources, so too are citizens saved valuable time by accessing the same information at their convenience on the ODP.

Citizens are not the only constituents self-serving their information needs through the ODP. Employees of governmental entities also use the ODP to discover and access their own data to complete their work. Using the ODP to host internal open data for data sharing makes government more efficient by reducing acquisition costs, redundancy and overhead. In addition, researchers from higher education and public education likewise have access to the open data that they need to complete projects and initiatives that support findings for the overall community.

2.3 The Open Data Lifecycle

The Open Data Lifecycle describes the process by which open data is selected, prepared, published, and consumed by the public. As the purpose of open data is to allow citizens free and open access to publicly held information, every part of the Open Data Lifecycle works with that end goal in mind within a user-centered design approach. When selecting which data to publish on the ODP, for example, consider first data that will receive the most public information requests by constituents. Knowing how data will ultimately be consumed likewise drives decisions about how to best prepare data for citizens.

It is important to note that publishing data is not simply making raw data available in an electronic format. Publishing also involves creating supporting documentation, metadata, and visualizations to give constituents a greater understanding of what the data represents. At the end of the Open Data Life Cycle, citizens consume open data and create outcomes, such as measuring the performance of government programs or creating new products and services.

Figure 1: Open Data Life Cycle

Tip: When selecting which data to publish to the Open Data Portal, consider first the information needs of end users in terms of fulfilling public information requests.

3. Open Data Portal

3.1 Background

The Texas Open Data Portal (ODP), located at https://data.texas.gov/ is the State's official website to access Texas open data. Both the ODP and the CDP instances are 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 data sharing platform.
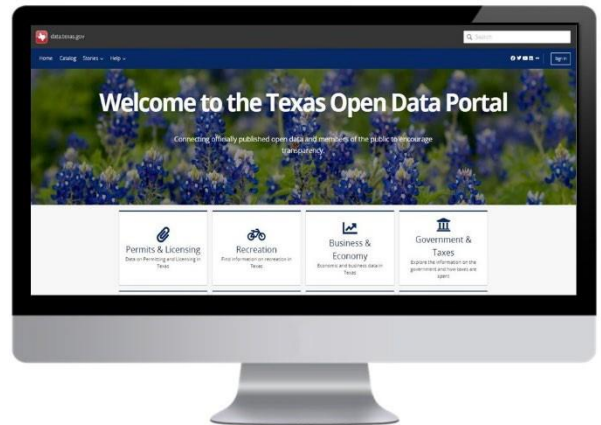
3.2 Roles and Responsibilities

DIR Open Data Portal Team

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

- granting access and publishing rights to agencies,
- managing the dataset inventory,
- providing guidance on open data governance,
- serving as a liaison between agency editors and the open data vendor-partner, and
- facilitating open data review/design sessions with agencies to highlight best practices of open data selection, preparation, and publishing.

The Open Data Portal Team is also responsible for assisting agencies in the set-up and design of CDP instances, including best practices for sharing private data.

Note: Although the Open Data Portal Team is willing to offer guidance and assistance, they do not edit, delete, or otherwise alter the data assets of publishing agencies.

Open Data Portal Users Group (ODPUG)

To further facilitate collaboration and best practice sharing among ODP users, DIR re-established the Open Data Portal Users Group (ODPUG) in 2020. The group is comprised of ODP users that have been designated by their respective agencies to lead the development of their open data management programs. The ODPUG meets regularly to discuss how to make improvements to the ODP, exchange success stories related to open data information sharing, and work with the Office of the Chief Data Officer (OCDO) on open data policies and related issues.

Socrata

Socrata, a division of Tyler Technologies, is the vendor under contract with DIR that provides the off-the-shelf open data platform used for the ODP and CDP instances.

Socrata is responsible for:

- accommodating a range of structured dataset sizes and copies of previously published Texas reports and maps,
- providing a support process to log issues, track status, and request system enhancements,
- implementing multiple level security, including encryption of data in transit and at rest, and controls that provide assurance of availability, and
- participating in technical guidance and design sessions around data use cases.

Agencies

Agencies are the owners of all data uploaded to the ODP and CDP, as outlined in the Open Data Portal Terms & Conditions available on the **Shared Technology Services portal**. As such, agencies are responsible for the content of the data published on the ODP, ensuring that data does not include content that either is or contains:

- objectionable,
- unlawful,
- in violation of a third party's copyright or intellectual rights,
- personally identifiable or private/confidential information,
- viruses or harmful files, or
- inaccurate or incomplete.

> ⚠️ **Warning:** Agencies should only publish a **copy** of their data. The ODP and CDP are not the system of record of an agency's data nor repositories.

4. Data Governance

4.1 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, apply to sensitive information, or can determine what type of information is available to the public.

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 public interest in the public information it houses.

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,
- the roles and responsibilities of open data stakeholders and publishers,
- an open data inventory,
- an approval/vetting process for publishing open data to the portal, and
- a process for regular auditing and improvement.

However, the formality of the charter or policies will vary by agency. Examples can be found on the OCDO SharePoint site in the documents of the Texas Open Data Portal section. Contact the ODP Administrator for access to this site.

Open Data Stewards

Once an eligible agency has completed the onboarding process, it must identify its Open Data Stewards. Each steward is responsible for managing the publication of open data for their agency in their program area. Depending on the size and number of data assets of the agency, the number of stewards can range from one to ten or more employees.

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

Based on DIR's security recommendations, each steward with Editor rights should have a unique ODP log-in linked to their government entity assigned email address. It is not recommended that data publishers share ODP IDs or passwords, as this increases the risk of security breaches.

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, and
- working with agency leadership in reporting and analyzing open data consumption.

Agency leadership from both business and IT functions should collaborate to determine which employees are 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
- Web Administrator
- Computer Operations Technician/Specialist
- Planner
- Systems Analyst
- Research Specialist
- Business Analyst
- Statistician/Economist
- Programmer
- Program Specialist/Manager
- Information Specialist

This list is not exhaustive but serves as a guideline for agencies. Depending on the level of technical complexity of open data publishing that the agency chooses to pursue, the skill level of its Open Data Stewards may vary.

It is critical that agencies understand their roles and responsibilities in publishing to the ODP. As *Section 3.2* outlines, agencies are solely responsible for the content, accuracy, and maintenance of their ODP data assets, and bear all legal responsibility in the case of a security breach. Therefore, it is key for agencies (and especially users with data publishing rights) to fully understand the characteristics of sensitive, protected, or confidential data that is protected by law. Examples of sensitive, confidential, or protected data are outlined in *Section 2: Open Data Overview*.

Warning: Agencies are responsible for implementing and following a data governance model that protects sensitive and confidential data from getting into the hands of unauthorized users.

4.2 Creating an Open Data Inventory

What is an Open Data Inventory?

As described in *Section 4.1*, each agency participating in the ODP 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 vast amounts of available data require agencies to prioritize the publication of certain data that will hold more value for constituents and data consumers.

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 consider for publication in the future.

Tip: Maintaining a data asset inventory is a continual process. Best practices dictate data governance teams assess the open data inventory on a bi-annual basis ensuring changes and updates are reflected in assets and metadata.

Open Data Inventory and Metadata

Your agency's datasets should also include applicable metadata. **Metadata**, or 'data about the data', provides information about each dataset. Metadata plays a vital role in the indexing, discoverability and understanding of each dataset.

Although the ODP currently only requires the fields of Title, Category Tile, and Agency, it is beneficial for agencies to collect more detailed information about each data asset when conducting an open data inventory. The more metadata that can be documented, the easier it is to find and understand the context surrounding published data.

Other than giving basic context of each data asset, the inventory should also provide information about:
- file type,
- size,
- quality,
- business owner and/or data steward(s),
- update frequency, and
- security/privacy constraints of the data.

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

"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, Former Assistant Commissioner, Texas Department of Agriculture

5. Data Preparation

5.1 Data Assessment and Selection

As your agency begins to build its Open Data Inventory, you must decide which data assets offer the most value to data consumers.

Approaching data selection from a user-centered design model is key to the success of an open data program. Select data for its value, quality, completeness, and appropriateness according to information needs. Remember, constituents can be internal or external to the agency.

Guiding Questions for Data Assessment

The following questions will help identify and prioritize open data to create an Open Data Inventory and prepare for publication.

- What information is frequently requested by the public, via the Public Information Request process?
- What data does the agency or program areas use for key performance indicators (either State Auditor's Office or internal measurements)?
- What data populates monthly, quarterly, or annual agency reporting requirements?
- What information is reported to local, state, or federal agencies?
- What information is shared internally across agency departments or divisions?
- What information is shared with other agencies?
- What information is shared with external stakeholders?
- What kinds of open data are 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),
- the ability to automate publication after the initial dataset is made public, and
- accessibility to data source systems to facilitate automation.

Defining a High-Value Dataset

As previously mentioned, SB 475 directs each Texas state agency with more than 150 full-time employees (FTEs) to post a minimum of three (3) high value datasets to the Open Data Portal.

What does the term 'high-value dataset' mean?

Section 2054.1265 of the Texas Government Code (TGC) defines a high-value dataset as information that can be used to

- increase state agency accountability and responsiveness,
- improve public knowledge of the agency and its operations,
- further the core mission of the agency,
- create economic opportunity, or
- respond to need and demand as identified through public consultation."

Warning: TGC Section 2054.1265 specifically states that a high-value dataset "**does not include** information that is confidential or protected from disclosure under state or federal law."

5.2 Data Quality

For data to be useful, it must be reliable, trustworthy and of high quality. Prior to publishing data on the ODP, data stewards should first complete several steps to ensure a high quality, interoperable and usable product for data consumers.

The core elements of data quality are:
- Completeness
- Uniqueness
- Timeliness
- Validity
- Accuracy
- Consistency

Data Quality Checkpoints

By considering the following data quality checkpoints, organizations maximize the discoverability, analysis, reuse and understanding of their open data assets.
- Datasets should be in a **machine-readable format** that allows for the use and reuse of data. Comma Separated Values (CSV) is the standard format for publication.
- Use **raw data** instead of calculated columns to make data more flexible and consumable in a variety of ways, including visualizations.
- Use **vertical orientation of data** rather than horizontal to make datasets more human readable and easier to create visualizations.
- Use **column names** that reflect natural language instead of the source data system.
- Ensure that data within the same field is **consistent** in terms of data type, format and expected values (e.g. avoid storing both text and numeric values in a numeric field).
- Use the **appropriate data types** that reflect how the data will be used. For example, numbers stored as text do not allow for calculations.
- Include **meaningful data points** instead of having blank cells or a value of "N/A" to allow for accurate calculations, filtering, or visualizations.
- If coded fields are included, add a **descriptive column** of the text equivalent for readability or include a data dictionary with the dataset, as a link or attachment.
- Exclude **symbols** such as currency, parenthesis, percentage symbols or commas from numeric data. Negative numbers should be indicated using the dash "-" symbol. Numeric columns can be formatted during the publication process to display currency, percentages, commas, and more.
- Numbers may include a varying number of **decimal places**. However, do not use commas or parenthesis. Decimal precision can be specified by column during the publication process.
- Where **leading zeroes** are meaningful (e.g., identification numbers or zip codes), store data as text to ensure that values are not truncated.
- Store **dates** in the MM/DD/YYYY format to make historical analysis easier. Date and time columns can be formatted in a variety of ways on the platform during the publication process.
- Remove **trailing spaces** so that operations which require an exact match are not affected.

Specific data quality checkpoints should be considered when adding a **geolocated column** to a dataset during the publication process for later use in map visualizations.

- Leave the address, city, state, and zip code in **separate columns**.
- **PO Boxes** are generally unable to be mapped.
- The platform utilizes **MapQuest t**o geocode. Addresses may be checked in MapQuest if they do not appear as expected.

5.3 Data Privacy

Data privacy is the part of data security concerned with data handling ethics. While data security protects data from internal and external attackers, data privacy governs how sensitive, confidential, or protected data is shared, collected, stored, and regulated. *Section 2.1* gives examples of data that is excluded from open data assets. This section outlines best practices in identifying sensitive, confidential, or protected data when determining what data should and should not be published on the ODP.

Guiding Questions for Data Privacy

- Does publishing the data violate any laws, rules, or regulations such as CJIS, HIPAA or FERPA?
- What are the intended and unintended effects of publishing the data?
- If the data does not contain personal identifiable information, can it still be combined with other public information to reveal sensitive details?
- Does aggregated and anonymized data have the potential to compromise internal processes, such as arrest records that expose where police concentrated their efforts?
- Does your organization own and have the right to publish the data? Was the data collected and compiled by a third party?

Tip: Always consult with your General Counsel regarding the rules and regulations specific to your agency's data publishing policies and practices.

"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

6. Publishing Open Data

6.1 Creating Metadata

A critical first step in publishing data on the ODP is creating metadata that describes the dataset and columns. Providing quality metadata is important to users because it creates access points to find, gives context to understand, and clarifies meaning to interpret the data. This metadata is added during the publishing process and is important to have identified prior to publishing.

Dataset Level Metadata

Dataset level metadata describes the dataset as a whole and helps users discover datasets in the ODP's data catalog. Dataset level metadata gives general information about the title, subject, and source of the data. At this time, only the dataset Title, Agency and Category Tile are required. However, agencies are encouraged to utilize the additional metadata fields to make datasets more discoverable and understandable.

Metadata Field	Justification	Guidelines
Title (Required)	Title helps discover and select datasets as well as differentiate between similar datasets.	Human-readable name of the dataset inplain language that includes enough detail to facilitate searching. Free of acronyms.
Description	Description helps discover and understand the content of datasets as well as differentiate between similar datasets.	Human-readable description with sufficient detail to enable the non- technical user to quickly understand whether the dataset is relevant to their search. Can include hyperlinks to additional information.
Row label	Row labels helps determine the appropriateness of the dataset for the user's needs.	Human-readable description of whateach row in the dataset represents.
Category	Not used on the Texas Open Data Portal	This field should be left as "-- No category --"
Tags / Keywords	Tags provide a means to include technical language, secondary categories, and acronyms.	Generally single or key words which help users discover the data and include terms used by technical andnon-technical users.
Public License Type	A license reduces legal uncertainty for data consumers or users about what rights they have to use and share your data.	Current list of licenses offered on the ODP platform can be found **HERE**. Consult your legal team for the best option to use or to leave unspecified as "-- No license --."

Data Provided By	Users can search by this field. It narrows down the data publisher within an agency for use statistics.	Department or organization that provided data. Must be consistent to utilize in agency metrics.
Source Link	Only if applicable.	If applicable, the link to the source ofthe data.
Contact Email Address	Users can ask questions about the dataset via a form, which is sent to this email address for further assistance.	Who understands what the dataset includes and can answer questions about it? This email address is not displayed to the data consumer.
Attachments	Attachments may provide extra context to the dataset.	Examples include terms of use or data dictionary. Commonly a PDF or Word document.
Agency (Required)	Users can browse the data catalog and filter datasets by the publishing agency.	Choose the agency responsible forpublishing the data from the dropdown menu.
Category Tile (Required)	Category provides a distinct navigation method and groups similar datasets together regardless of source.	Choose the category of the dataset from the dropdown list of possible values. If a dataset can fall into multiple categories, select the one which is most significant.

Table 1: Dataset Metadata Fields

Dataset Column Level Metadata

Like a data dictionary, column level metadata helps data consumers better interpret and use datasets on the ODP by providing context and meaning to the data. By creating quality column level metadata, agencies allow their constituents greater access to the information they seek.

Metadata Field	Guidelines
Column Name (Required)	Human-readable name of a column in plain language that is free from jargon and acronyms. Autogenerated during the publication process based on the source's column names but can be edited as needed.
Column Description	Human-readable description with sufficient detail to enable the user to interpret data in a column.

Table 2: Column Metadata Fields

Tip: When creating metadata, consider what terms make the most sense to end users. Using language that most closely matches user search terms and expectations will help users better find, identify, select, and obtain the information they seek.

Creating Metadata for Other Asset Types

Metadata is equally important for other data asset types, allowing them to appear in the catalog and search results. **Story pages, data lenses, maps and charts** have metadata that is edited after the asset has been published.

For story pages, go to the list of assets in the profile of the story page owner or co-owner and locate the story page to be edited. To the right of the story title under the "Actions" column, click on the three dots and choose "Edit metadata."

For data lens, maps and charts, simply append "/edit_metadata" to the asset URL to access the metadata editing tool.

The metadata fields are similar to the dataset metadata fields. Although Title, Description, Tags/Keywords, Agency and Category Tile have "*****" before the field name, only Title is required. It is best practice to complete these fields so that the asset appears in the catalog and searches.

The following fields are not used for these asset types in the ODP and should be left blank:

- Category
- Row Class,
- Subject Column,
- Resource Name,
- Row Identifier,
- Upload Image, and
- New Attachment.

Learn more about metadata best practices HERE.

6.2 Manual Publishing

Manual publishing is performed by users with publishing rights, called data editors, and is the simplest way to upload a dataset. Using this method, an editor can upload an individual file. Possible formats include .csv, .xls, .xlsx, .tsv, or shapefiles. At the most basic level, dataset publishing follows this method after clicking the "Create" button and choosing "Dataset" from the dropdown:

- Enter the name of the dataset.
- Add data by selecting and uploading the file.
- In the resulting data table preview, review and select a column to be the **row identifier**, if the dataset has one.
- Correct the column data types and column level formatting. Make any necessary data transformations. Information on these optional but helpful functions are available HERE.
- Correct any errors in the data, such as text in a number column.

- Create a georeferenced column if the dataset contains location information. This ensures the dataset can be visualized as a map later.
- Reorder or hide columns, if necessary.
- The dataset can be saved as a draft to come back to later in the "My Assets" section of the owner's profile page. Otherwise, the dataset can be published privately.
- Add other ODP users to collaborate or view the dataset for internal vetting prior to publishing publicly. Follow the agency's open data governance policies and procedures.

Step by step directions for uploading, reviewing, and formatting are available HERE.

6.3 Automated Publishing

Automated publishing is the process of updating data programmatically, rather than manually. The advantage of using automated publishing is that updates can be scheduled without the delay of relying on an individual to manually refresh data. This is especially helpful for data that is updated frequently. This section highlights automated publishing using DataSync and Socrata API functionality.

DataSync
The **Socrata DataSync** tool is a free and powerful publishing tool that lets users schedule and automate data updates and upload large data files to the ODP. The program requires Java version 8 or higher and can be downloaded for free from Socrata's DataSync site on GitHub.
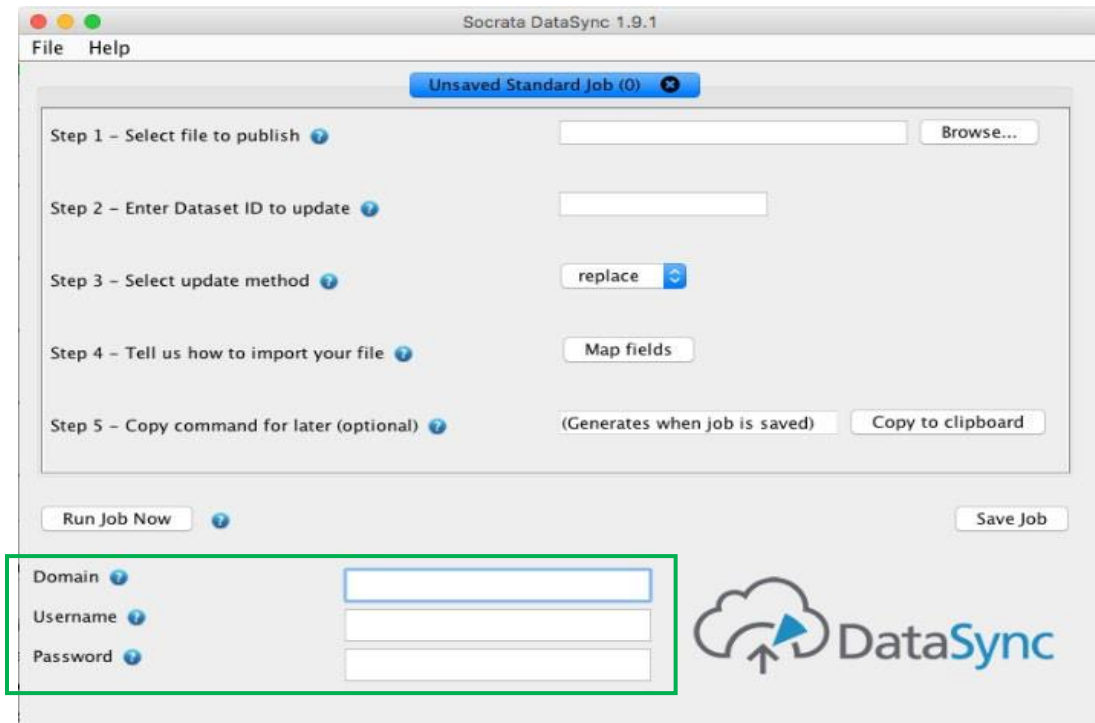

Figure 2: DataSync Automated Publishing Set-up

The roled user setting up the automation should have an "Editor" role on the ODP and either own or co-own the ODP dataset.

After downloading DataSync, the basic outline for setting up a standard job is below. Detailed step-by-step instructions are available **HERE**. Additional resources for each step are linked below where applicable.

1. Set up and save the **DataSync preferences**. This is essential if the agency's network has a firewall or proxy server set-up. Common considerations are detailed **HERE**.
2. Complete the Authentication Details Section (green outline in *Figure 2*), entering the following information:
   > **Domain:** https://data.texas.gov
   > **Username/Password:** Socrata Open Data Portal login credentials.
3. Select the file to publish (.csv/.tsv files only).
4. Enter Dataset ID to update (the 8-character code separated by a dash at the end of a dataset's URL). This is also known as the "4 by 4."
5. Select update method (option to replace, upsert, append, or delete). Note that upsert and delete require the dataset to have a **Row Identifier**.
6. Indicate how to import the file and map fields. Select options within the **Advanced Settings** as necessary.
7. Run the DataSync job by selecting "Run Job Now." Then, click "Save Job" to save the command (optional, used to run from the command line or to schedule future jobs).

Socrata API Resources

API stands for Application Programming Interface, which allows the ODP to talk to other products or services. There are multiple APIs that open up data and functionality to developers, including automated publishing. For more information and technical resources, visit Socrata's Developer site at **https://dev.socrata.com/**.

> **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).

Socrata APIs utilize a query language known as **SoQL**, the Socrata Query Language, which is similar to SQL. Each uploaded dataset has a built-in **Socrata Open Data API (SODA)** endpoint, which offers data filter and querying capabilities using SoQL. This **endpoint** is simply a unique URL representing the dataset and is used to interact with data resources.

Other important APIs include:
- **Discovery API:** explore the metadata of data assets via read-only access,
- **Metadata API:** update asset metadata programmatically, and
- **Dataset Management API:** upload, update and publish data programmatically – including column metadata.

A few situations where an agency might consider using an API include:

- updating data on the ODP and then updating the same data on an agency's website as a separate process,
- updating a dataset at regular intervals, or
- due to limited time and/or resources, updating open data on a consistent, manual basis is not feasible.

Automated data publishing should be implemented with the full compliance of an agency's data governance policy and processes.

6.4 Dataset Backups

A **Dataset Snapshot** allows you to keep an up-to-date version of your dataset on hand within the platform itself, in case of a failed update or any other need to revert to a previous copy. These snapshots are not discoverable in the catalog.

Although there are conditions in which a backup is created automatically, it is good practice to manually backup your dataset on a regular interval or if there will be a major change to the dataset. As a reminder, any data on the ODP or CDP should be a copy, not the source record of the data.

6.5 Publishing Process

Whether you are using the manual or automated method to publish datasets to the ODP, the publishing process should generally follow the cycle illustrated in *Figure 3*.
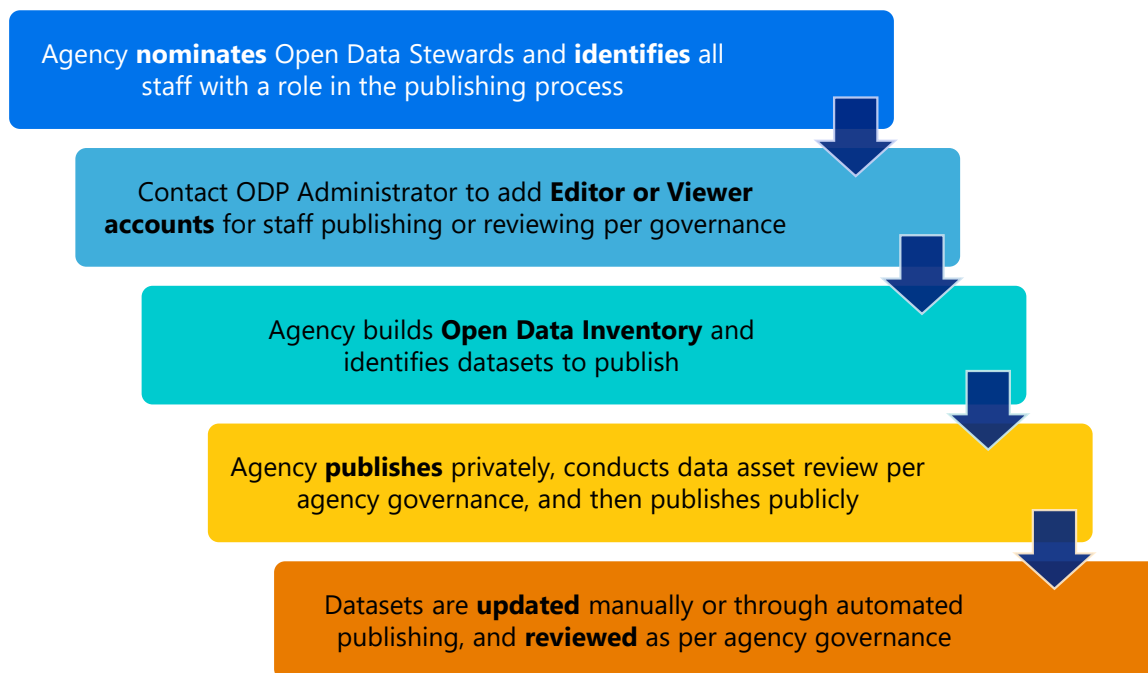
Agency **nominates** Open Data Stewards and **identifies** all staff with a role in the publishing process

Contact ODP Administrator to add **Editor or Viewer accounts** for staff publishing or reviewing per governance

Agency builds **Open Data Inventory** and identifies datasets to publish

Agency **publishes** privately, conducts data asset review per agency governance, and then publishes publicly

Datasets are **updated** manually or through automated publishing, and **reviewed** as per agency governance

Figure 3: Publishing Process

Private vs. Public Data Publishing

The ODP has a necessary function for data editors – the ability to draft and publish a data asset privately, before making it available publicly. When a data editor creates a new asset, it remains in draft mode until "Publish" is selected. Likewise, a data asset can be published in a "Private" mode. In both scenarios, only data asset owners and those users who have been granted access by the owner can view or make changes to the asset.

Asset owners can share permissions with other users on the domain to allow for vetting of data assets per agency governance. Data assets can be reviewed before public publication, ensuring that data quality, accuracy, privacy and/or security concerns are addressed. More details about the sharing of data assets and the asset view level options are HERE. The individual asset level permissions are different than the ODP domain roles of Editor and Viewer assigned by the ODP Administrator. This allows a greater level of access control to individual assets.

7. Data Consumption

In addition to the data publishing methods described above, the ODP and CDP provide data analytics, visualization and communication capabilities for data editors and viewers. Each feature described below updates automatically as the underlying data is refreshed. This section provides general guidance in using these functions, but more detailed information, tutorials, and videos are linked in each section.

An important clarification: when a user with the Editor role publicly publishes data assets, they are marked as "Official," meaning they have been created by a publishing agency.

The public data consumer can create and save filtered views, maps, or charts of official datasets, which will be marked as "Community" assets. However, they cannot publish or edit any original data.

7.1 Dataset Analytics

Data consumers have the option of working with the data directly on the platform or with their own data analytics software off the platform. Data can be manipulated for analysis in a variety of ways including sorting, filtering, conditional formatting, and aggregation. As mentioned, filtered views can be saved by anyone with a Socrata account, whether public data consumer or agency data publisher.

The OData endpoint is generated automatically for all datasets. This endpoint easily connects a dataset to external third-party tools like Tableau, Power BI and Excel for data analytics off the platform. The benefit is that as dataset updates occur on the ODP, the data can be refreshed in the analytics tool via this connection to remain current.

7.2 Visualizations

Visualizations available on the ODP can be customized in a variety of ways. Visit Socrata's Data Visualization and Exploration documentation for step-by-step instructions on creating and customizing charts and basic or layered maps.

A special feature of mapping is the ability to use Point Aggregation. Both heat maps and region maps are available options. Please contact the ODP Administrator for assistance if the custom boundary you wish to use for a region map is listed under "Requires Processing" or if you have a custom boundary (spatial lens) to be added to the ODP.
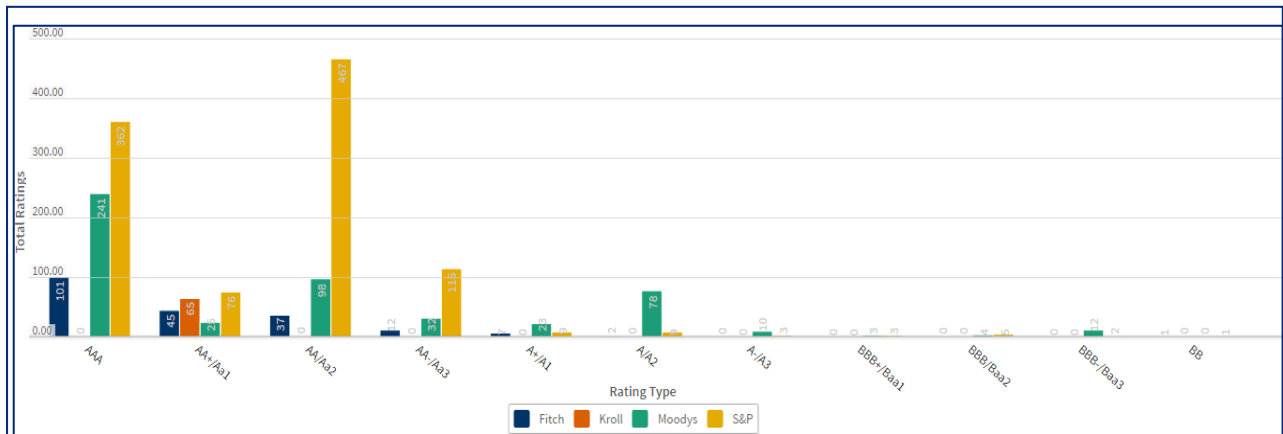
Figure 4: Clustered Bar Chart
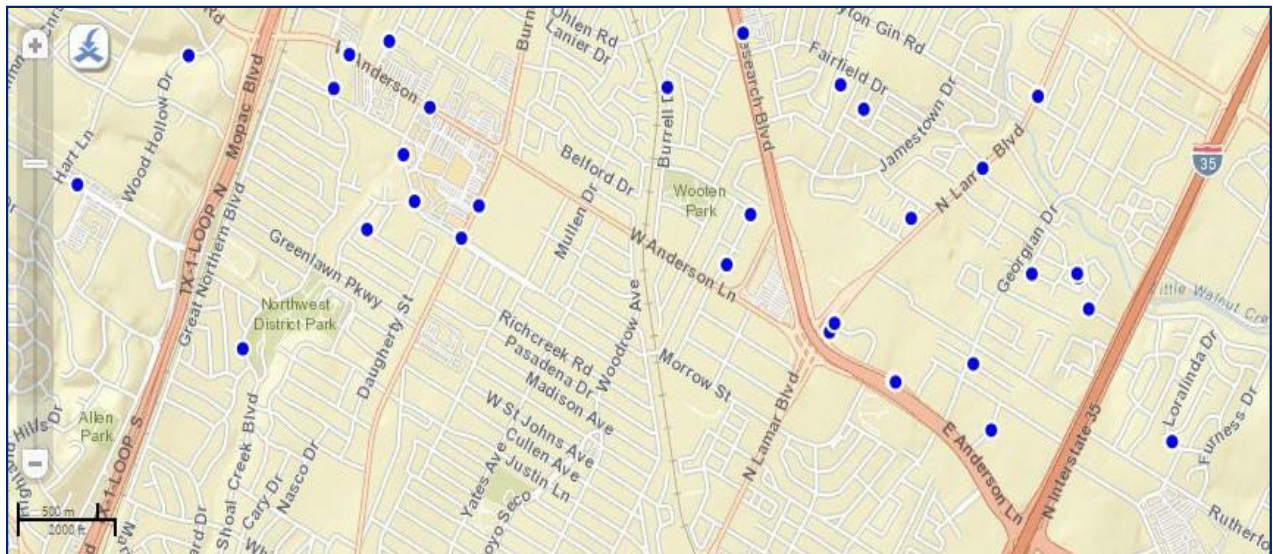
Figure 5: Point Map
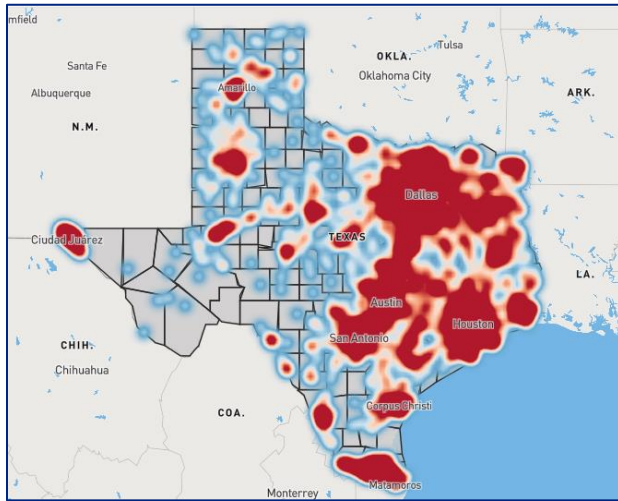
Figure 6: Layered Heat Map
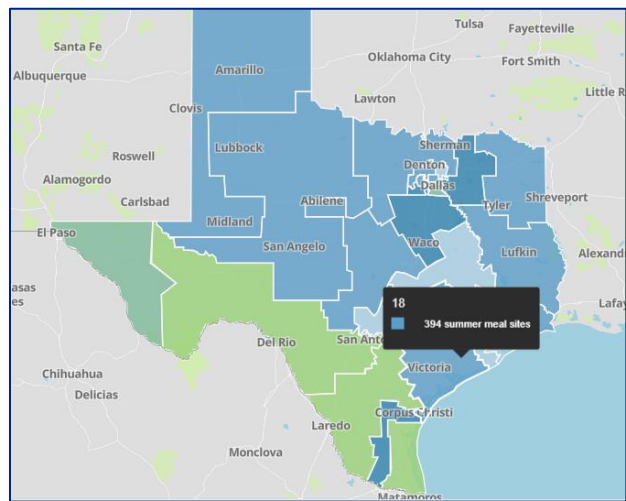
Figure 7: Regional Map, aggregated by Senate District boundaries

7.3 Data Lens

A [Data Lens](#) allows ODP editors to create data visualizations exploring one dataset in a variety of ways. The data consumer can easily interact with the data through visual filters depicted in [cards](#), which graphically represent columns of the dataset. Types of cards include:

- Columns, for categorical data,
- Timelines, aggregated by date,
- Maps, both point and choropleth,
- Distribution, for numerical data, and
- Search, providing a look-up feature.

Figure 8: Data Lens with active filtering

7.4 Stories

Story pages, also known as Perspectives, are multimedia assets that combine text, visualizations, and images to tell a story about an agency's data and programs. While individual datasets allow access to the raw data, Stories provide greater context and meaning as to what the data represents and why it is important to data consumers. Stories are outreach tools that highlight and bring recognition to the important work agencies do for the public.

Once a Story is public, the Story Tile can be embedded on any webpage on the internet, including another Story page. The Story Tiles include the title, description, the top image, and a link to view the full Story.

Figure 9 is an example of a Story published on the ODP that includes an embedded map, a performance measure, explanatory text to provide context to an agency program, and external links. Because Stories are dynamic, the embedded content such as visualizations and performance measures are refreshed in real-time as the underlying ODP datasets are updated.

Figure 9: Story Page

Using simple drag and drop content blocks, data editors have the following options for adding and editing text and rich media content in Stories:

- **Presentation Ideas:** text content blocks, including table of contents and multi-columns.
- **Global Filter Bar:** allow users to filter charts, maps, and tables that are powered by the same dataset.
- **Visualizations:** create or insert an existing chart, map or measure.
- **Tables:** create a hierarchical table based on a Socrata dataset.
- **Story Tile:** insert a preview of another story.
- **YouTube:** add a video from a YouTube link.
- **Image:** add an image from a computer file or from Socrata's Getty images library.
- **HTML Embed:** insert an embed code to include external web content, such as HTML coding snippets, Tableau or Power BI visualizations and more.

> **Tip:** The ODP has an **Interactive Element Demo and Library** with customizable HTML code to create interaction in your Story Pages. Ask the ODP Administrator for guidance.

7.4 Notifications

ODP users have two methods to stay up to date on platform actions and datasets that they are interested in following. Notifications can either be site-driven or dataset-driven. Examples of site actions a user may choose in their notification settings include when any new asset is published, or any asset is deleted. Users can also watch a dataset and be notified of any changes or updates to that particular asset.

8. Checklist for Getting Started on the Open Data Portal

It is the role of the DIR Administrator of the Open Data Portal (ODP) to onboard new agencies as well as support existing publishing agencies in their open data journey. The following checklist is used to onboard and ensure continued success in data sharing.

The ODP Administrator is a resource and supports users in each of the following steps:

- ☐ Prospective agency agrees to the **terms and conditions** of the ODP.
- ☐ Agency identifies **data stewards** who create a **data inventory** of potential open data assets and may also act as data editors.
- ☐ Agency submits the names of **data stewards and/or data editors** to the ODP Administrator, who will register authorized staff on the Socrata platform as roled ODP users.
- ☐ Data editors follow the Socrata Open Data Portal Education Plan (*Section 9)* to learn how to use the platform.

- Agency open data governance team **selects and prepares data for publication** on the ODP. The DIR Administrator of the ODP is available as a resource to assist with identifying data for ODP publication.
- Data editors practice publishing and other skills on the **Sandbox** instance of the ODP at https://texas-test.demo.socrata.com/login.
- Data editors work with data stewards to **create metadata and publish** open data to the ODP.
- Data stewards and data editors attend the quarterly **Open Data Portal Users Group (ODPUG)** to network, share and learn from other users.
- ODP users and members of the open data governance team **attend trainings, webinars and conferences** hosted by the Office of the Chief Data Officer, including the annual **Data Forum**.

An important note: The Sandbox environment is shared with agencies involved in the learning process. It is used to practice publishing and other skills and the site homepage/catalog is not accessible by the public. The Sandbox is not linked to the Texas Open Data Portal (https://data.texas.gov) and assets cannot be transferred from the Sandbox to the ODP (or vice versa). The ODP is for staging assets in the private mode (not yet available to the public) so they can be vetted and approved appropriately, prior to publishing to the public.

The Sandbox is not linked anywhere for the general public. However, roled users can see all publicly published Sandbox assets, and links to those public assets can be shared outside of the Sandbox. With a direct link, that publicly published data asset is viewable by anyone. This is an **example: https://texas-test.demo.socrata.com/dataset/Stacey-Test/54ky-4gku**.

You can publish privately on the Sandbox and add permissions for others to access the data asset, if you want to share with other users. You can also change your data asset back to private at any point if you published publicly while practicing. This is an example of a dataset on the Sandbox published privately: **https://texas-test.demo.socrata.com/dataset/Metrics-Test/s8bd-tph2**. It should redirect you to the Sandbox log-in page, or if you are logged into the Sandbox as a roled user, you should not be able to access this asset.
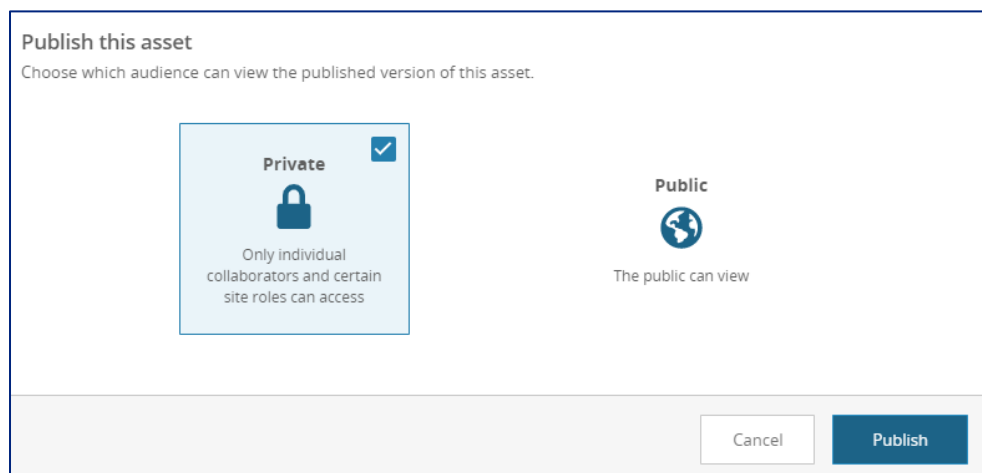

Figure 10: Publishing choices in the ODP and Sandbox

9. Open Data Portal Education Plan

Socrata has a library of on-demand and live web courses designed to help data stewards and editors learn the platform. To set up a free learning account, visit Socrata's Education site at http://learn.socrata.com/?access_code=texas_dir using the code texas_dir.

9.1 Recommended Courses for New Users

1. The Story of Open Data (15 minutes on-demand)
2. Introduction to the World of Data on Socrata (13 minutes on-demand)
3. Prepare Your Data (25 minutes on-demand)
4. Clean and Tidy Data (60 minutes, live or on-demand)
5. Create and Manage Your Dataset (90 minutes live or 30 minutes on-demand)
6. Explore Data with Charts (60 minutes, live or on-demand)
7. Map Your Data (60 minutes, live or on-demand)
8. Tell a Story with Perspectives (16 minutes on-demand)

9.2 Additional Advanced Courses

1. Developing Your Data Governance Policies and Standards (30 minutes, on-demand)
2. Map Out Your Digital Center of Excellence Strategy (10 minutes, on-demand)
3. Create Performance Measures (90 minutes, live or on-demand)
4. Transform Data with Socrata Query Language (60 minutes, live or on-demand)
5. Create and Update Your Dataset with a Socrata Publishing API (16 minutes on-demand)
6. Automate Data Updates with a Python Template (26 minutes on-demand)
7. Data Analysis Tools and Connections (60 minutes, live or on-demand)

10. Closed Data Portal

While the ODP is the official central repository for all publicly accessible open datasets for the state of Texas, the Closed Data Portal (CDP) is a private data sharing environment intended to host Sensitive, confidential, or protected data. Use cases for the CDP include vertical data sharing between departments of the same agency or horizontal data sharing between two or more agencies. The Socrata platform is FedRAMP Moderate authorized to operate and therefore CDP instances can host data containing PII and PHI. There is no connection or data sharing between the ODP and CDP, nor between unique CDP instances.

Agencies may request their own CDP instance and receive assistance from the ODP Administrator to set-up the CDP to their specific needs. After initial set-up, the CDP instance is administered by the agency, and the ODP Administrator remains available for support, but no longer has access to the CDP instance. Access to any CDP instance is granted only by invitation from the administering agency and user accounts may be disabled at any time.

Please note: The CDP cannot host CJIS data. Any CDP that will host HIPAA data requires additional agreements for compliance. The ODP Administrator at DIR will assist with that process.

Feature	Open Data Portal	Closed Data Portal
Free of Charge to Agencies	X	X
Create, Manage, & Publish Datasets	X	X
Visualize Data with Charts & Maps	X	X
Data Automation (DataSync and API)	X	X
APIs for Editing, Metadata, & Discovery	X	X
Private Data Sharing with Specific Audiences		X
Guest Role (share with external stakeholders)		X
Custom Dataset Alerts		X
Socrata Gateway		X

Table 3: ODP and CDP Comparison

Please note: to access links below about CDP features and capabilities, create a Tyler Technologies Socrata Support Account at https://support.socrata.com/hc/en-us. Click Sign In at top right and then the Sign Up link. This is separate from the ODP or Sandbox log-in.

11. CDP Administration

Administrators of a Closed Data Portal instance manage three areas of the site: data, content, and settings. To access the Administration panel for these functions, click on the gear icon and then choose "Administration." Managing data includes viewing audit trails of all activity on the site as well as site usage analytics. Site content management includes configuring the appearance of the site, and an approval process prior to any public publication. Finally, managing site settings includes creating and managing users, roles and permissions as well as creating a site-wide metadata scheme used to categorize all data assets. Best practice is to assign the role of administrator to at least two staff members to allow for coverage.

Please note: to access links below, create a Tyler Technologies Socrata Support Account at https://support.socrata.com/hc/en-us. Click Sign In at top right and then the Sign Up link.
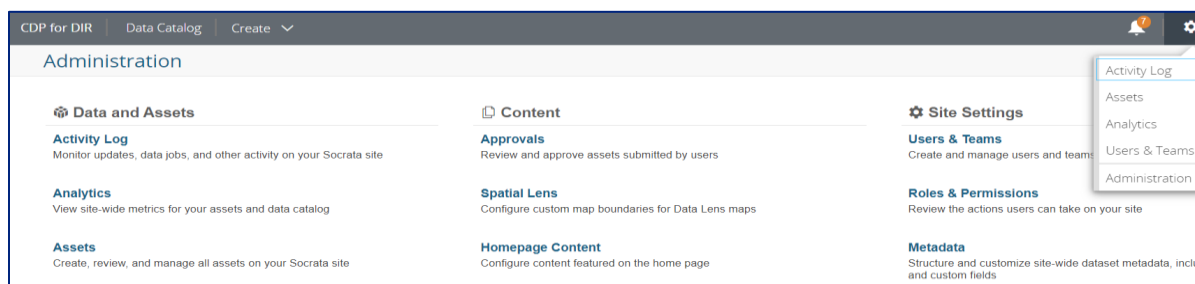

Figure 11: CDP Administration Panel

11.1 CDP Data Management

Auditing

CDP Administrators have access to the Activity Log for the CDP instance. Accessible from the Administration panel, the Activity Log tracks changes to CDP instance, including the publishing or deletion of an asset, asset permission changes, data update failures. This additional insight into the CDP's activity is essential for audit capabilities.

The Asset List is also a useful auditing tool for CDP Administrators, displaying all data assets on the site including those in draft mode or published privately. The Asset List is also available from the Administration panel by choosing "Assets."

Site Analytics

Site Analytics are available to CDP Administrators through several system datasets. These datasets are automatically updated daily, published privately and only accessible by those users with an Administrator role on that instance. These datasets can be filtered, sorted, and visualized just like any other dataset to provide valuable insight into the CDP's usage.

- **Asset Inventory:** complete inventory of all assets on the CDP instance used to track the performance of data publishing or conduct metadata management.
- **Asset Access**: asset interaction by user type (views, downloads, and API reads).
- **User Authentications**: hourly user log-in authentications, including user ID
- **Catalog Search Terms**: time-based count of words and phrases entered into catalog search bars by users.
- **Referrers**: previous webpage a user was on when following a link to the site. This dataset may only appear if there are one or more publicly published assets.

11.2 CDP Content Management

Site Configuration

CDP Administrators are able to customize the look and the functionality of the CDP instance through the Administration panel.

Homepage Content is managed through content management datasets. These datasets power items on the homepage including section headings, banner image, category tiles and featured stories tiles.

Site Appearance provides configuration of the header, footer and navigation of the CDP instance. Color is also customizable by adding branded palettes for visualization colors.

Approvals

The need to publicly publish on the CDP is rare due to the sensitive nature of the data, which makes the Approval Process feature especially beneficial. CDP Administrators can moderate the publishing of assets using custom Approver roles. These approvers must review and approve data assets prior to their public publication. The implementation of this two-step process diminishes the inadvertent release of sensitive, confidential, or protected data by publishing publicly. Data is still able to be published internally to all roled users on the CDP instance in one step.

11.3 CDP Site Settings Management

User Roles

In order to share or view a CDP data asset, that individual must have a role on the CDP site. The CDP Administrator manages all site access, including adding or disabling users through the Administration panel. There are three **default roles** available on the CDP:

- **Administrator**
- **Internal User –** this role can view any asset that is internally published, as well as be added as a viewer or a collaborator to privately published assets. Internal users have publishing rights on the CDP instance.
- **Guest –** this role is for those external to your organization that may require viewing access. Guests can be added to Teams and allowed Collaborator or Viewer rights. Guests are not able to publish new datasets and are limited in their ability to publish views and visualizations of data assets that they can access, if added as a collaborator.

Custom Roles

A major benefit of the CDP is the ability to create <u>Custom User Roles</u>. When setting up a custom user role, the CDP Administrator assigns rights on a granular level. Use the <u>Rights Glossary</u> to determine the access parameters for custom roles.

Teams

The <u>Teams</u> feature is an efficient way to provide access control if there are specific users that will need to be granted the same level of access to data assets. Instead of assigning access user by user to each data asset, users can be added to a Team. The Team can then be assigned access to the necessary assets.

Metadata

Although the Socrata platform provides a <u>standard metadata schema</u>, CDP Administrators have the ability to add custom metadata fields to ensure each asset contains a complete and standardized information required for the agency's needs. The new <u>Metadata Template</u> interface offers a user-friendly way to customize Categories, make a metadata field multiple choice, and audit metadata for all assets. For best practices on metadata management and examples, visit <u>HERE</u>.

11.4 CDP-Specific Features

Alerts

<u>Alerts</u> are dataset-driven and customizable based on the user's parameters. Alerts may be time-based and occur on a schedule, scan the entire dataset or just scan the new rows to see if a particular condition is met.

Gateway

The CDP provides another method for publishing and automating datasets, <u>Socrata Gateway</u>. Gateway offers a variety of plug-ins to connect the user agent, a Java client set up on a server with access to the source system hosting your data, to Socrata's back-end systems. The full list of Gateway plug-ins can be accessed <u>HERE</u>. Monitoring of scheduled Gateway automation can easily

be selected in a user's notifications, quickly alerting to automation fails. Agent Logs and Agent Events are tracked for auditing purposes and troubleshooting. Similar to asset access, Gateway access is governed by a user's role on the CDP instance.

12. CDP Training and Resources

In addition to the Recommended Courses #3-8 in the Socrata Open Data Education Plan (see *Section 9*), CDP Administrators will benefit from the following.

Note: the CDP is referred to as the Socrata Connected Government Cloud (SCGC) in Socrata's help articles and education courses.

12.1 Additional Recommended Courses for CDP Administrators

1. Socrata Connected Government Cloud Site Administration (2 hours live or 45 minutes on-demand)
2. Shape & Join Data with SoQL Query Editor (60 minutes, live or on-demand)
3. Check Domain Activity with Site Analytics (25 minutes on-demand)

12.2 Knowledge Articles

- Reference Guide to Sharing Data on SCGC
- Creating a Filtered View on the CDP

13. Additional Resources and Links

13.1 Support

- **Socrata Client Center:** https://support.socrata.com/hc/en-us
- **Socrata Support Ticket Submission:** https://support.socrata.com/hc/en-us/requests/new
- **Socrata Developer Portal:** https://dev.socrata.com/

13.2 External Open Data Portals

- **Federal Open Data Portal:** https://www.data.gov/
- **City of Austin Open Data Portal:** https://data.austintexas.gov/
- **City of Dallas Open Data Portal:** https://www.dallasopendata.com/

14. Glossary of Terms

API – stands for Application Program Interface. API is programming code that allows for two distinct applications to communicate with each other in an automated way.

Closed Data Portal – a private data sharing environment intended to host private or sensitive data. Unlike the Open Data Portal (ODP), access to data in the Closed Data Portal (CDP) is by invitation only.

Data – information that has been translated into binary components that computers can move, store and process.

Dataset – a collection of data, usually in a tabular format, that is published on the ODP.

DataSync – the Socrata proprietary tool that has been developed to automate the data publishing process for ODP users.

Data Steward – the person or persons responsible for management and oversight of an agency's data. Data stewards provide a vital role in ensuring data quality, availability, and consistency on behalf of the data owner.

Editor – an ODP user with publishing rights.

FedRAMP – Federal Risk and Authorization Management Program established in 2011 to provide a cost-effective, risk-based approach for the adoption and use of cloud services by the federal government. Authorization to operate is designated at Low, Moderate or High impact levels.

Machine-readable – data in a standard format that can be easily recognized and processed by computers.

Metadata – 'data about the data' that provides the contextual understanding about each dataset.

Open Data – data that refers to information that can be freely used, re-used, and redistributed by anyone.

Open Data Governance Framework – a 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 – 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 Data Management Framework (TDMF) – the methodology of core principles for developing individual agency Enterprise Information Management programs.

Texas Public Information Act – the policy of the State that, unless expressed otherwise by law, guarantees citizens complete information about the affairs of government and the official acts of public officials and employees.

15. Appendix: Open Data Legislation in Texas

Texas Public Information Act – Title 5, Open Government, Ethics, Chapter 552 – Public Information (September 1, 1993)
Set forth the policy of the State that, unless expressed otherwise by law, guarantees citizens complete information about the affairs of government and the official acts of public officials and employees.

Open Data – SB 701 82(R) (September 1, 2011)
Required state agencies to post high-value datasets on a generally accessible agency website. High-value datasets include, but are not limited to, data that is critical to the financial and programmatic function of state agencies.

Open Data – SB 279 83(R) (September 1, 2013)
Required state agencies to send the Department of Information Resources (DIR) a description and link to their high value datasets.

Statewide Data Coordinator – HB 1912 84(R) (September 1, 2015)
Created position to collaboratively develop data policies, standards, and best practices and to improve data governance and integrity statewide.

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.

Chief Data Officer and Texas Open Data Portal – SB 819 86(R) (September 1, 2019)
Established the position of Chief Data Officer (CDO) for the State of Texas and established the Open Data Portal (ODP) as the official central repository for publicly accessible electronic data.

Data Management Officer and ODP Publishing Requirements – SB 475 87(R) (June 14, 2021)
Required agencies with more than 150 full-time employees to designate a Data Management Officer, who has the responsibility of publishing or assigning a designee to publish at least three high-value datasets on the Texas Open Data Portal.