



# Getting Your Data Into Socrata

---

The Wide World of Ingress Methods

# The Socrata Client Center

[Your starting point for all things Socrata](#)

Product News & Updates, Educational Resources, [How-To Videos](#) and more

One article to rule them all: [Overview of Ingress Methods](#)

# Ingress Methods

Manual upload

Import from URL

Adding an external reference

Building extract, transform, and load pipelines

Using APIs or other software development tools

Connecting with other catalogs

Accepting federated assets from another Socrata domain

# Manual Methods

# Manually Upload a Dataset Using the Socrata Data Management Experience

Best for one-off, ad-hoc use cases or for data that is not frequently updated where automation would provide limited benefit

Low barrier to entry for non-technical users

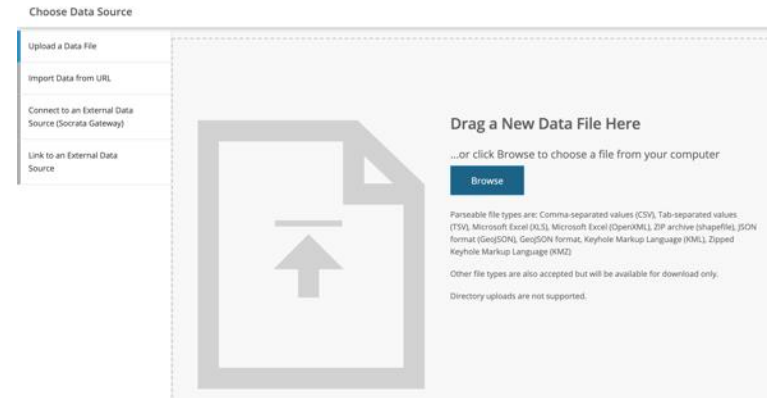
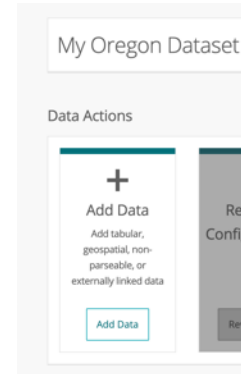
Socrata allows for drag and drop or selection of a file to upload data

Data is parsed and a [graphical interface guides](#) users through the process, such as verifying data types and defining metadata

Advanced functions like transforms and geocoding are optional

Automation workflows can be generated for Python or any other automation pathway can be used to update an existing dataset

Limited scheduling and alerting



# Don't See the Create Button?

The ability to Create datasets is permission controlled

If you do not see the Create button, please work with a system administrator to adjust your role

# Import Data From a URL

Import from any external, parseable location such as a publicly available internet URL

Parseable formats are currently defined as .csv, .tsv, .xls, .xlsx, .zip (shapefile), .json (GeoJSON), .kml, .kmz

Can support non-parseable formats as attachments that are discoverable in your data catalog

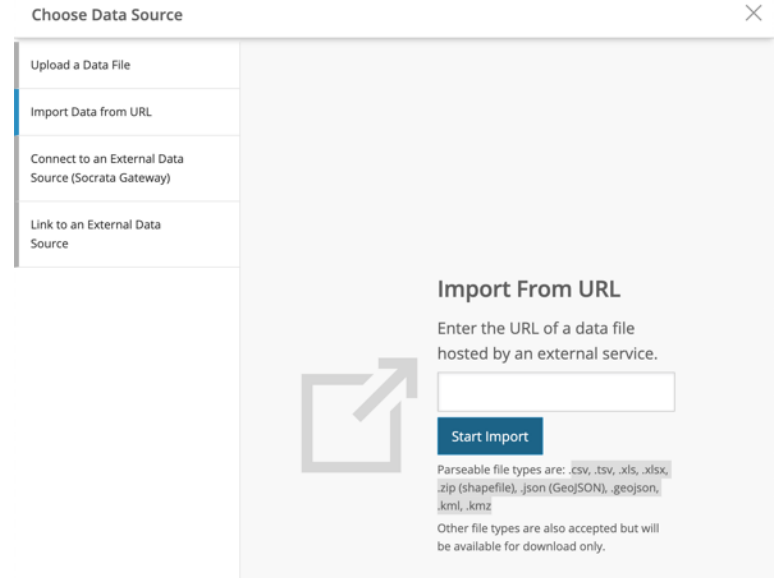
OnDemand or Schedulable

No need for scripting, APIs, or other software development

Great for third-party data or legacy published data

Datasets are only as current as published external location

A great method to ingest from other Socrata sites, the Census, or Data.gov

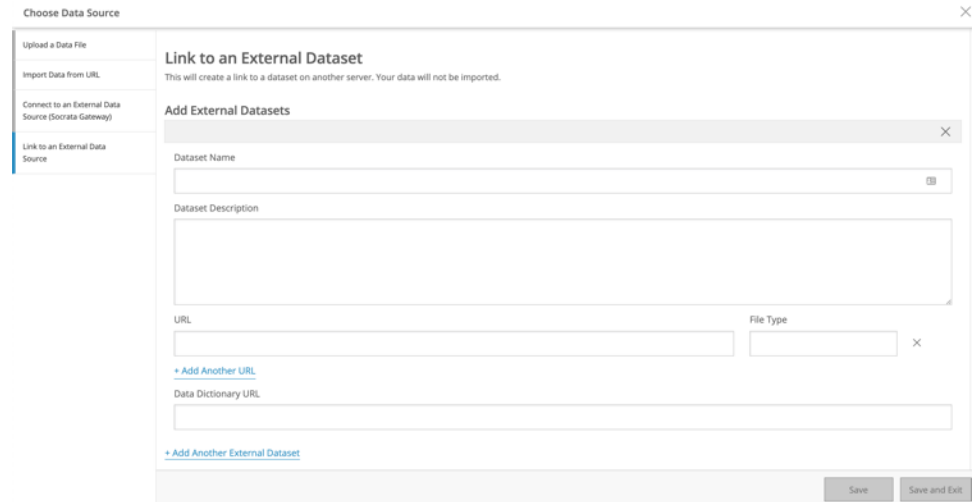


The screenshot shows a 'Choose Data Source' dialog box with a sidebar menu on the left and a main content area on the right. The sidebar menu includes the following options: 'Upload a Data File', 'Import Data from URL' (which is highlighted with a blue bar), 'Connect to an External Data Source (Socrata Gateway)', and 'Link to an External Data Source'. The main content area is titled 'Import From URL' and contains the text 'Enter the URL of a data file hosted by an external service.' followed by an empty text input field. Below the input field is a blue 'Start Import' button. At the bottom of the main content area, there is a list of parseable file types: .csv, .tsv, .xls, .xlsx, .zip (shapefile), .json (GeoJSON), .geojson, .kml, .kmz. A note below this list states: 'Other file types are also accepted but will be available for download only.'

# Link to an External Dataset

External Datasets are indexed as part of your catalog, but the data is not ingested into your domain; it exists as a reference and when used, the user is navigated to the external location

A name and description can be maintained, as well as the file type, URL, and if provided, a data dictionary



The screenshot shows a web interface titled "Choose Data Source" with a sidebar on the left containing four options: "Upload a Data File", "Import Data from URL", "Connect to an External Data Source (Socrata Gateway)", and "Link to an External Data Source". The "Link to an External Data Source" option is selected. The main content area is titled "Link to an External Dataset" and includes a warning: "This will create a link to a dataset on another server. Your data will not be imported." Below this is a section titled "Add External Datasets" with a close button (X). The form contains the following fields: "Dataset Name" (text input), "Dataset Description" (text area), "URL" (text input), "File Type" (text input with a close button), "+ Add Another URL" (link), "Data Dictionary URL" (text input), and "+ Add Another External Dataset" (link). At the bottom right are "Save" and "Save and Exit" buttons.



# Ingress Considerations

---

When planning to create a dataset, it is important to consider many things to help settle on the right pathway

Sample Considerations:

- How often does the data need to be updated?
- Does the data need to be manipulated on load?
- Can the source system of the data be integrated with directly?
- Is a development effort needed to access and load the data?
- What is the size of the data being uploaded?
- And more

# More Advanced Methods and Automation

# More Advanced Ingress Methods

Socrata provide many fully-featured APIs which can be used directly or have been implemented by many third-party extract, transform, and load (ETL) tools such as FME or Talend

API use or Third-Party integrations are managed off the Socrata platform

Socrata provides its own ETL tool, [the Socrata Gateway](#)

The Socrata Gateway [Plugins](#) can integrate with a variety of existing applications, database standards, government data sources, and more

# Gateway Agent

Setting up a Gateway agent requires a system administrator and potentially information technology resources

Requires a thin-Java client to be installed on your infrastructure, a server or virtual machine

Specific Java versions and environmental configuration is needed

The client is unique to your domain and downloaded from it

Once configured, Agents can be used to ingest data by non-system administrators

Local or cloud ingestion available from the client

# Gateway Plugins

[Gateway plugins](#) are containerized integrations

The Agent calls plugins, which are developed by Socrata for dozen of applications, including common business applications like Excel to common government applications like Accela to government information stores like the Census to common DB technologies like MySQL

Plugins are installed per agent and code signed

With a Gateway Agent configured, plugins are a low barrier form of automation; no coding is required

Allows for notifications and [flexible scheduling](#)

For stable integrations, not ad-hoc use cases

# SAFE FME or Other ETL

[FME](#) provides an engine to extract, transform, and load from a variety of source systems and to a variety of source systems

FME has [read and write integration with Socrata](#)

Other tools, Pentaho or SSIS, can be employed in concert with the Publishing API

# Socrata APIs

Socrata has several API enabled publication methods; the [Publishing API](#) is the current standard, but legacy [APIs](#) do exist

Use of the API requires independent, custom development and authentication

Users can publish data, update data, and set a schedule of their own control

API endpoints are agnostic of the clients preferred programming language and execution infrastructure, local or cloud

Requires technical acumen

# Socrata Developers

[API Docs, Publication Guides, and other Application Development resources](#)



# Large Scale Ingress for the Admins: Catalog Connectors and Federation

# Catalog Connectors

Your Socrata data catalog can ingest or provide external references to existing published data

Connect [Esri](#) or [data.json](#) catalogs

Connect all assets or a subset of assets

Schedule daily (Esri and data.json) or monthly updates (Esri)

data.json connector must conform to [DCAT-US Schema v1.1 \(Project Open Data Metadata Schema\)](#)

Setting up connectors is an admin level function so talk to your admins if you have an integration need

# Federation

Your Socrata data catalog can [federate](#) assets from another Socrata catalog or federate data to another Socrata catalog

Common integration between levels of State and Local Government

For example, Austin, TX federates data to the State of Texas and the State of Texas accepts the federated data from Austin, TX

Federation is a system admin level function and requires an agreement between federating parties

# Which Is Right For You?

*Unsure of which method to choose?*

Review the [Pros and Cons](#) of each

Review the ingress considerations

Talk to your data steward

Talk to your Socrata [Client Success Manager](#) or [Support Specialist](#)

Employ [Socrata Coaching Hours](#) to explore more advanced scenarios

Empowering people who serve the public<sup>®</sup>



**tyler**  
technologies

# Appendix

# Additional Ingress Considerations

What ingress methods are you currently using to ingest data into the platform?

Are you satisfied with the existing workflow(s) and if not, why are you considering moving away from them?

What are some of the reasons that you would like to automate your data?

Will data be automated on frequent basis: daily, on-demand or subtask?

Will data be automated on a rolling basis?

What are the format and source systems of the data?

What other data source systems exist within your organization that you'd like to automate?

What size are the datasets (tables, views and queries) that you are looking to automate?

Will you consistently add or update the database with new tables, views and queries?

What existing ETL software tools are you leveraging to extract data from those source systems?