# Manually Importing A Dataset

Montana Data Portal

## What types of files can be imported? (For Importing a dataset step by step instructions, navigate to page 7)

**Import a Data File:**

1. .csv (csv output from an excel file) **(Recommended format)**
2. .xls (Excel 97 and later)
3. .xlsx (2007 and later)
4. .tsv

Please Note: .csv is the optimal file format for upload. If you experience an error while importing a .xls or .xlsx file, try saving as a .csv to eliminate Excel formatting that can sometimes cause errors.

**Upload a Non-Data File:**

1. Any file type

**Import Geospatial Data:**

1. .zip (shapefiles; the following files are required: .shp, .shx, .dbf, .prj)
2. .kml (We do not support multiple features in a single layer. Technical terminology: we do not support heterogeneous children inside a multi-geometry tag)
3. .kmz (this a zipped .kml and follows the same import rules as above)

**Link to External Data:**

1. URL

**Connect with an ESRI Map Layer:**

1. URL to the RESTful endpoint for an individual map layer on an ArcGIS Server version 10.0 or above.
2. Currently our platform supports Web Mercator projections via ESRI Extensions and will not convert other projections types.

## Import Detected and Supported Field and Column Types

There are many other data types that we support but you'll have to either convert your columns after import or enter the data manually.

The following types are supported:

### Plain Text

UTF-8 encoded text (generally alphanumeric text). Our system assumes that there is no text formatting.

### Formatted Text

UTF-8 encoded text that may contain html.

### Numbers, Money and Percent

For numbers we directly use Java's BigDecimal parsing.

For negative numbers, the format should **not** contain any commas. *-10000* will be read correctly but *-10,000* will not.

A percent can be either a number preceded or followed by a percent (%) sign or just a number. Percentages aren't in the range 0.0 to 1.0 like they are in Excel. A percentage input of "*42.0*" is idiomatically "*42.0%*".

Money can be either a number preceded with a dollar sign ($ -- more currency symbols soon) or just a number. For negative monetary values, either a negative sign or a set of parentheses are acceptable: e.g. *$-42.21*, *($42.21)*, *-$42.21* or *(42.21)*.

### Dates & Time / Date & Time (with timezone)

Dates are parsed by default in the American/Pacific (PST) timezone. You can explicitly specify a timezone by using the supported ISO 8601 subset. A 'Z' character is UTC, otherwise the offset is *[+-]HH:mm*.

For inputs that don't specify a time of date, the resulting time is undefined. In other words, don't rely on it being anything consistent.

The accepted input formats are:

### Supported ISO 8601 Subset

- *yyyy-MM-dd['T']HH:mm:ssZ* (e.g. "*1920-01-22T00:00:00Z*", "*1920-01-22T00:00:00-10:00*", or "*1920-01-22 00:00:00Z*")
- *yyyy-MM-dd['T']HH:mm:ss* (e.g. "*1920-01-22T00:00:00*" or "*1920-01-22 00:00:00*")
- *yyyy-MM-dd['T']HH:mm* (e.g. "*1920-01-22T00:00*")
- *yyyy-MM-d*d (e.g. "*1920-01-22*")


### Supported non-ISO Dates

For dates other than the ISO subset we accept a date, optionally followed by a time, i.e. *(date)[ (time)]*

Non-ISO dates are always parsed in the American date format locale (i.e. month, day, year). Months and days can be either single or double digit and may or may not be led with a '0'. Years can be either four digits (preferred) or two. If a year is two digits it will be assumed to be between 1951 and 2050: i.e. *1/2/75* would be January 2nd 1975, but *1/2/49* would be January 2nd 2049.

The accepted input formats are:

- *MMM d, yyyy* (e.g. "*Jan 4, 1982*")
- *MMM d, yy* (e.g. "*Jan 4, 82*")
- *MMMM d, yyyy* (e.g. "*January 4, 1982*")
- *MMMM d, yy* (e.g. "*January 4, 82*")
- *M-d-yyyy* (e.g. "*1-4-1982*")
- *M/d/yyyy* (e.g. "*1/4/1982*")
- *M.d.yyyy* (e.g. "*1.4.1982*")
- *M-d-yy* (e.g. "*1-4-82*")
- *M/d/yy* (e.g. "*1/4/82*")
- *M.d.yy* (e.g. "*1.4.82*")

## Location Columns

Location columns are a "composite" column that's created by appending multiple values together. We accept the following types of location data for geo-location:

### Example of Excel:

Example of the imported results, including what the data syntax looks like in Excel and when Imported.

| Excel Format Description | Excel Data Type | Excel Data Format | Imported Data Format | Imported Field View Style | Map Compatible |
|---|---|---|---|---|---|
| Combined Address with No Quotes | Text | 125 N Roberts ST, Helena, MT, 59620 | 125 N Roberts ST Helena, MT 59620 (46.5854498380005°, -112.01650103799972°) | Address & Coordinates | Yes |
| Combined Address Quotes | Text | "125 N Roberts ST, Helena, MT, 59620" | "" (46.585438°, -112.015744°) | Address & Coordinates | Yes |
| Latitude & Longitude Combined No Quotes | Text | (46.585438-112.015744) | 46 585438 | Address & Coordinates | No |
| Latitude & Longitude Combined With Quotes | Text | "(46.585438, -112.015744)" | "(46.585438, -112.015744)" | Address & Coordinates | No |
| Separate Latitude & Longitude Columns | Text | Separate Columns | Helena, MT 59620 (46.585438°, -112.015744°) | Address & Coordinates | Yes |

### Example of CSV:

Example of the imported results, including what the data syntax looks like in CSV format and when imported.

| CSV Format Description | CSV Data Type | CSV Data Format (No headers, Common Delimited) | Imported Data Format | Imported Field View Style | Map Compatible |
|---|---|---|---|---|---|
| Combined Address with No Quotes | Text | 125 N Roberts ST, Helena, MT, 59620 | 46 585438 | Address & Coordinates | No |
| Combined Address Quotes | Text | "125 N Roberts ST, Helena, MT, 59620" | 125 N Roberts ST Helena, MT 59620 (46.5854498380005°, -112.01650103799972°) | Address & Coordinates | Yes |
| Latitude & Longitude Combined No Quotes | Text | (46.585438-112.015744) | 46 585438 | Address & Coordinates | No |
| Latitude & Longitude Combined With Quotes | Text | "(46.585438, -112.015744)" | "" (46.585438°, -112.015744°) | Address & Coordinates | Yes |
| Separate Latitude & Longitude Columns | Text | Separate Columns | Helena, MT 59620 (46.585438°, -112.015744°) | Address & Coordinates | Yes |

# Recommended Location Format:

1. Separate Columns for Street Address (US only). (In Excel and CSV as text fields)

| Full Street Address | City | State | Zip |
|---|---|---|---|
| 125 N Roberts ST | Helena | MT | 59620 |

2. Separate Latitude and Longitude columns. (In Excel and CSV as text fields)

| Latitude | Longitude |
|---|---|
| 46.585438 | -112.015744 |

Make sure that your values are in decimal degrees, and that you use "negative" longitude degrees for "degrees west".

## Website Links/URLs

URL's support two different input formats. Only three URL schemes are acceptable: ftp, http, and https. We use a custom regular expression to validate URLs. It should accept just about anything that you throw at it, but there's always a chance that it's missed something.

1. *<a href="http://www.socrata.com/">Socrata</a>*
2. *http://www.socrata.com/*

## Emails
Three different input formats are acceptable for emails.

1. *<a href="mailto:sam.gibson@socrata.com">Sam Gibson</a>*
2. *sam.gibson@socrata.com*
3. *Sam Gibson <sam.gibson@socrata.com>*

Nearly all emails should work, though technically for performance' sake we only support a subset of the RFC regex for emails. If there's a specific email or set of emails that's causing you a problem, please feel free to submit a support ticket and we'll fix it.

## Checkboxes
Valid false values:      Valid true values:

- *0*
- *f*
- *false*
- *n*
- *no*
- *off*

- *1*
- *t*
- *true*
- *y*
- *yes*
- *on*

## What data types are available after import?

There are some file types that we do not handle during import, but you can add to your dataset after you have imported the file.

### Phone

- Number. We do not do validation to confirm if it is a valid phone number or format.

### Multiple Choice

- You can pre-enter values that a user can select from a drop-down.

### Photo (Image)

- Accepted file formats: .jpg, .png, .gif

### Document

- We accept any file type

## How to Manually Import a Dataset:

Select the **Import a Data File** option.

**Design from Scratch:** Choose this if you don't have a data file to import yet. You will be able to define a dataset schema and input the data later online.

**Import a Data File:** Choose this if you already have a data file on your computer which you wish to import. There are four file types allowed - .csv, .xls, .xlsx and .tsv.

**Upload a Non-Data File:** If your data file is not in any of the four allowed formats, such as a PDF or an image file, you are still able to host it on the data portal via this option. Note that though the file which still be searchable, it cannot be interacted with as with the other four file types.

**Link to External Data:** Choose this is you want to link to data hosted on another site. This data will not be imported.

**Import Geospatial Data** and **Connect with an ESRI Map Layer** are detailed in Maps section under 'Visualization Tools'

**Create a new Dataset**

How would you like to create your new Dataset?

| Design from Scratch | Import a Data File | Connect with an ESRI Map Layer |
|---|---|---|
| Import Geospatial Data | Upload a Non-Data File | Link to External Data |
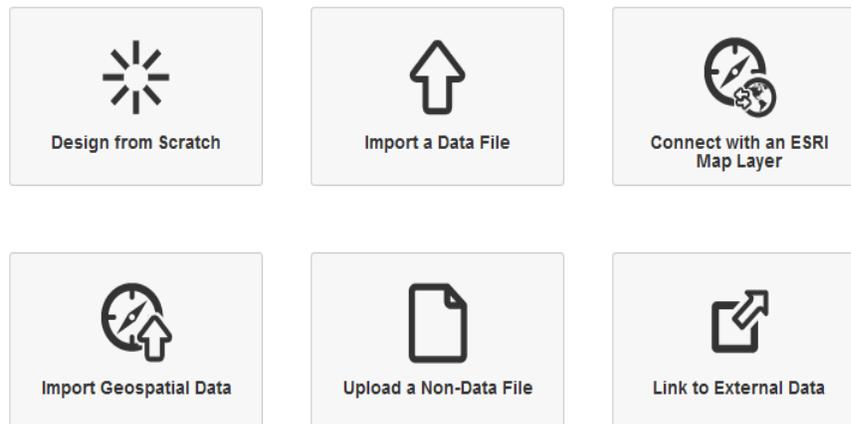
Cancel                                                         Previous    Next

Select where the File is stored.  For this example select '**On my computer**'

### Where is this file located?

| On my computer | On the Internet |
|---|---|

If you select this option, you will be prompted to select the file from your machine. Select Upload file button, navigate to the file location, and select the file

### Please choose a file to import

| No file selected yet. | Upload a file |
|---|---|

Supported formats are .csv, .tsv, .xls, and .xlsx.

*If you select 'On the internet':* *Choose this if your data lives on the internet in .csv, .xls, .xlsx or .tsv, and you have the HTTP(S) URL to it.*

*If you select this option, you will next be prompted to provide this URL:*

### Please indicate where your data file lives.

| What's the web URL of the file you'd like to import? | Import |
|---|---|

Supported file types are CSV, TSV, and Excel. Supported protocols are HTTP and HTTPS.

Manually Importing A Dataset

This stage allows you to review the columns of data which are being imported and the overall schema of the dataset, including the column ordering, column data type, and the number of rows which are headers. You are able to change the name of the columns, data type and source column. Select the next button.



**Name:** The name is text fields which are automatically taken from the first row of the file as the header, which will be the column header in the dataset. These fields can be edited if you want to change the name of the column during this stage. These column headers can be changed when the dataset is imported, as well.

**Data Type:** The system reads the first few rows of each column to make an educated guess on the data type of the column. If the guess is wrong, this can be corrected by selecting the appropriate type from the menu which appears when you click on the data type.

**Source columns:** These are the columns from the underlying dataset whose entries will populate the platform dataset under the 'Name' columns. Additional options include:

- adding a row
- deleting a row
- clearing all rows to start with an empty dataset
- resetting the schema configuration changes to the original

Under **Headers,** the system will by default read the first row as the header. You can use the 'Fewer Rows' and 'More Rows' to set the accurate number of rows which are enabled, if it isn't the first.

## The Metadata is descriptive data about the dataset.

Enclosed below is a description of what to enter into the Metadata form.  Please be sure to enter the required information reflected with a red star '*'.

| Field | Description |
|---|---|
| **Dataset Title\*:** | This is the name that will appear on the portal. |
| **Brief Description\*:** | Explanation about the data set. Description may include an abstract or detailed text. |
| **Primary and Secondary Contacts, and Primary Mobile Phone Number\*+:** | Identify Primary and Secondary Contact Names, email addresses and mobile phone numbers. These will be used as a reference to route questions and comments concerning the data set. |
| **Data Provided By\*:** | Department, Division, Agency or Organization |
| **Category\*:** | Identify the category in which the data set will appear on the portal (only one may be selected).  Select from the following:  Business, Drivers, Education and Training, Employment, Environment, Family Services, Financial, Health, Information and Referrals, Licenses, Permits, Politics, Property, Public Safety, Recreation, Shopping, Taxes, Tourism, Transportation, Vehicles, Voting |
| **Tags/Keywords:** | Searchable keywords, or data set tags; these are to help the customer find the data set. They also provide ways to find other data sets that may be similar. Separate tags with commas. |
| **Update Frequency:** | Frequency when changes and additions are made to the data set after the initial data set is released. For example, a data set can be updated daily, weekly, bi-monthly, monthly, quarterly, semi-annually, annually, as needed, etc. |
| **Dataset Accuracy:** | Comments about the accuracy of the data set, if applicable. |
| **Dataset Security+:** | A data set can be either public or private. If the data set is considered private (available only to those authorized), please list the Active Directory group containing those that may access this data.  (Applicable for state employees only.) |
| **Source Link:** | Sources related to the data set can be provided.  Include text explaining the source and then a link where the customer can find additional information. |
| **Additional Disclaimer:** | There is a standard disclaimer for the portal.  If there are more restrictive terms, or additional disclaimers needed for the data set, provide that information. |
| **Thumbnail Image Upload Image:** | A special image (icon) may be associated with each data set.  Please indicate yes/no as to whether an image is desired.  If yes, please provide the image through the SITSD Service Desk (as described at the top of this document). |
| **Attachments:** | Attachments (such as a .pdf document) may be associated with each data set. Please indicate yes/no as to whether there will be any attachments and the name of the file(s).  Attachments may be provided through the SITSD Service |

| Field | Description |
|---|---|
| | desk (as described at the top of this document). |
| **Comments⁺:** | List any other pertinent information or special instructions regarding this data set. These comments are for internal use only and will not be displayed on the portal. |
| **Contact Email⁺:** | Please input the dataportal@mt.gov email as the 'contact email'. The Data Portal Manager will coordinate and forward if necessary any emails received to the Agency Primary Contact. Otherwise, the dataset owner will receive all emails from the public and agencies through the 'About' tab and clicking on the contact dataset owner action. |

\* = Required Field
\+ = Information for internal use only, not displayed on the portal

After entering the required information, select the next button. *Please note that the default dataset permission is private.*

**Create a new Dataset**

Please describe "Demo Test Dataset"

* Dataset Title — Demo Test Dataset

Brief Description — Enter a description

Category — -- No category --

Tags / Keywords — Enter keywords
Enter one or more keywords separated by commas

**Licensing & Attribution**

Licensing — -- No License --
Choose licensing terms for people who wish to use this dataset.

Data Provided By — Individual or Organization

Source Link — Enter Web Address

**Attachments**

To upload attachments, first finish publishing this dataset. Then, click 🛈 About in the toolbar, and click Edit Metadata. You will then be able to upload attachments.

**Data Set Custodian Information**

* Primary Contact Name
This field has been marked private by your organization and is not publicly viewable

Primary Cell Phone Number

Secondary Contact Name
This field has been marked private by your organization and is not publicly viewable

**Other Information**

Update Frequency

Additional Disclaimer

Data Set Accuracy

Comments
This field has been marked private by your organization and is not publicly viewable

**Privacy & Security**

Privacy Settings — ○ My data is **public**: it can only be modified by me, but everyone can view it
● My data is **private**

Contact Email — Email Address
This address will not be displayed publicly. If left blank, it will default to your address

* Required Field

Cancel                                                   Previous   Next

Newly imported datasets are initially 'Working Copies', which means that the data columns can be edited, sorted, and filtered, if needed. Select Ok.  If no columns need updating, proceed to 'the menu options' header on the next page. Otherwise, scroll down to the 'Editing Columns' section for details.
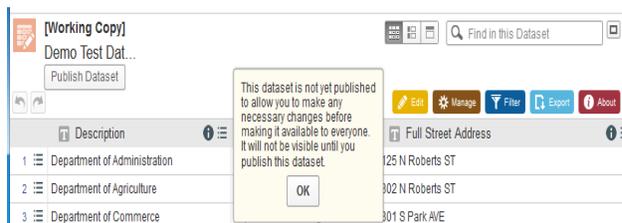


## Editing Columns:

To edit a column, select the menu button. , and Edit Column Properties.



An option list appears on the right side of the dataset.  The available options are to change the name of the column, include a description of the data in the column, change a dataset, data alignment, total a column, and defining the API Identifier (This field needs to be unique). Select update for changes and cancel for no changes.  Select the Update button when completed.
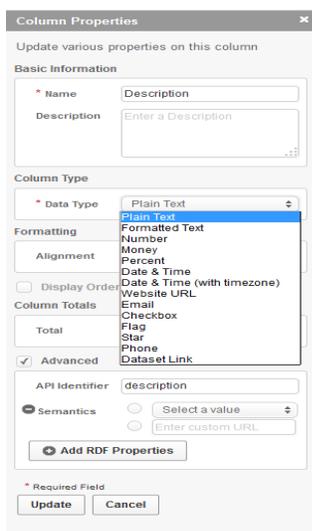
## The menu options will allow greater customization.

Edit   Manage   Filter   Export   About

The edit button   Edit

> Under the edit button, there are options for Manual Append & Replace Wizard, and adding an additional Column to your dataset.   Please see the 'Manually Appending and Replacing datasets' document for details on how to utilize this function.  Select Create if you are adding a column.

**Edit** ✕
Append and Replace ◀
**Add Column** ▼

Add a new column to your dataset

**Basic Information**

| * Name | Enter a Name |
| Description | Enter a Description |

**Column Type**

| * Data Type | Select a data type |

* Required Field

Create   Cancel

The manage button   Manage

> Under the Manage Ownership, you are able to assign ownership to someone else by entering their email address and clicking Apply.

**Manage** ✕
Ownership ▼

To transfer to another user, enter the email address or account name of the Socrata ID, then select the desired account from the dropdown list. You may also enter the user's profile URL if you know it. Only users with existing accounts may take ownership of datasets.

**Transfer ownership to**

Apply   Cancel

Delete ◀
Sharing ◀
Permissions ◀
Show & Hide Columns ◀
Column Order ◀

13

Under Delete you can delete a working copy.
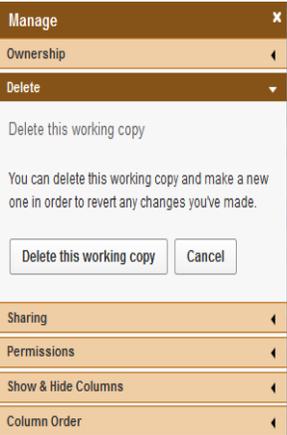


Under Sharing you can select additional users able to access the dataset by entering their email address.   You can select viewer, contributor, or owner access.  Select Email when completed.
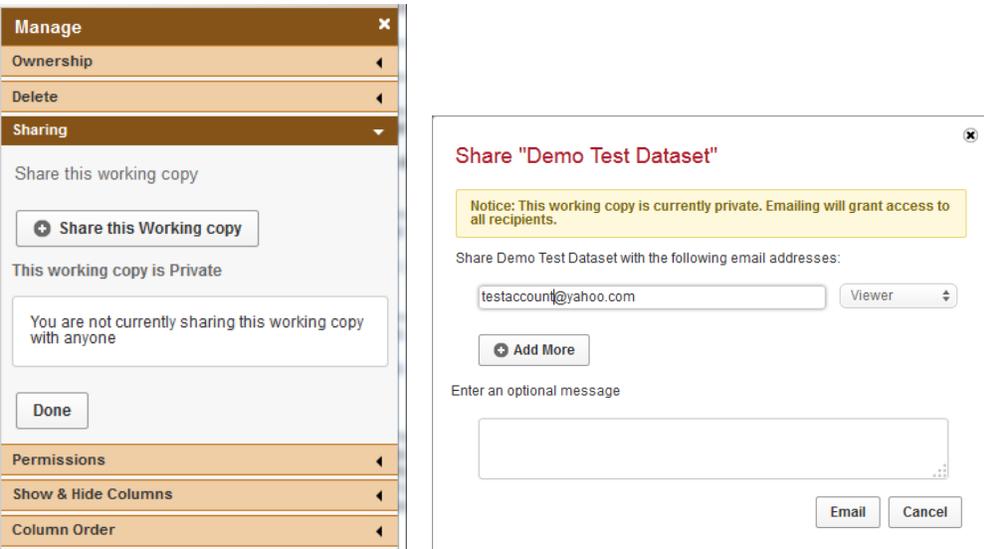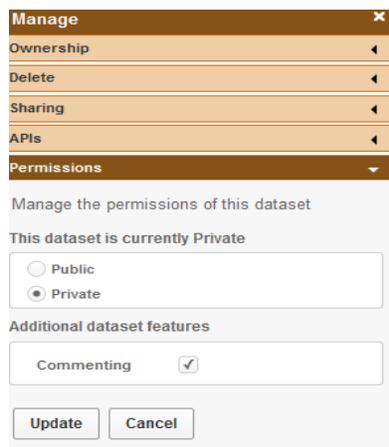


When the user that the dataset was shared to logs in and select the 'Shared To Me' header, they will see the dataset.

Manually Importing A Dataset
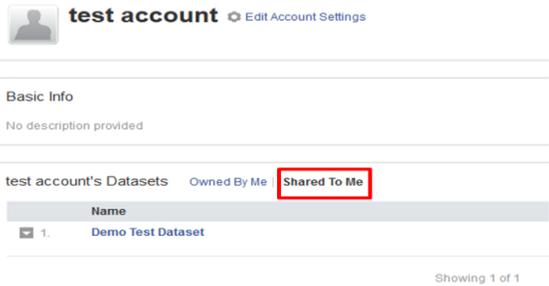
Under API's you can select to create documentation for an API by selecting 'Create an API'.
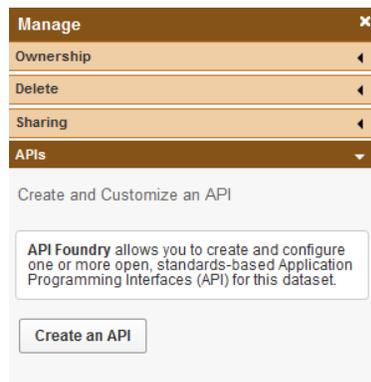


Under Permissions, you can select to change a dataset from private to public and enable or disable commenting on your dataset.



Under Show & Hide Columns and Column Order, there are options to select which columns users see in your dataset and the order of the columns.

Manually Importing A Dataset
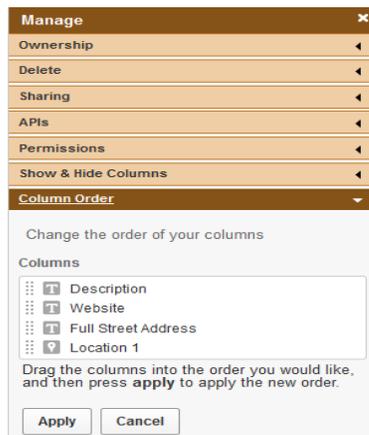
The filter button

Under the Sort & Roll-up section, options for sorting by a specific column or columns are available.  Select the sort button, select the column from the list, select the sort type, add column for additional columns, and select apply.
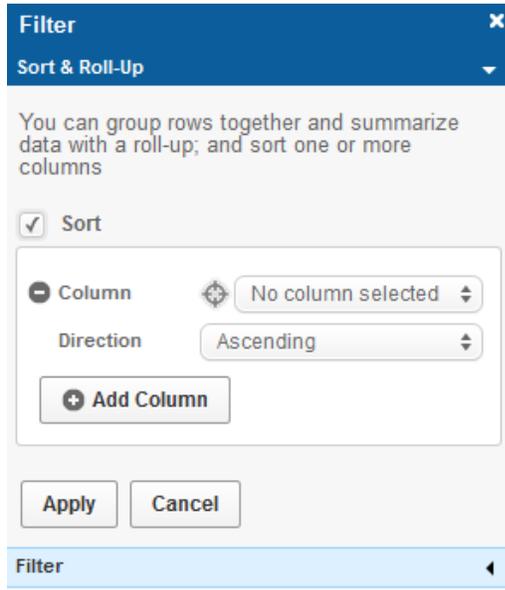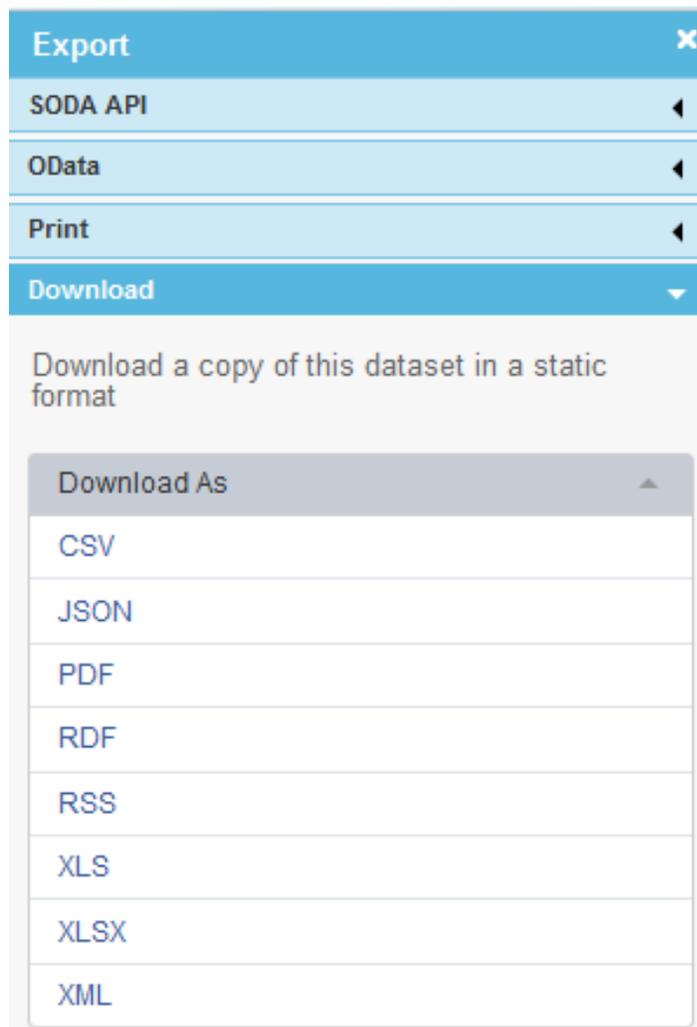
Under the Filter section, there are options for the dataset to be filtered based on criteria. Select add a new Filter Condition, select the field (in the example below Description is selected, select the operator for the filter (recommendation is to use contains). To designate the filter value enter it in the field (in the example below Administration) Unchecking the box returns all values in the dataset.

Manually Importing A Dataset
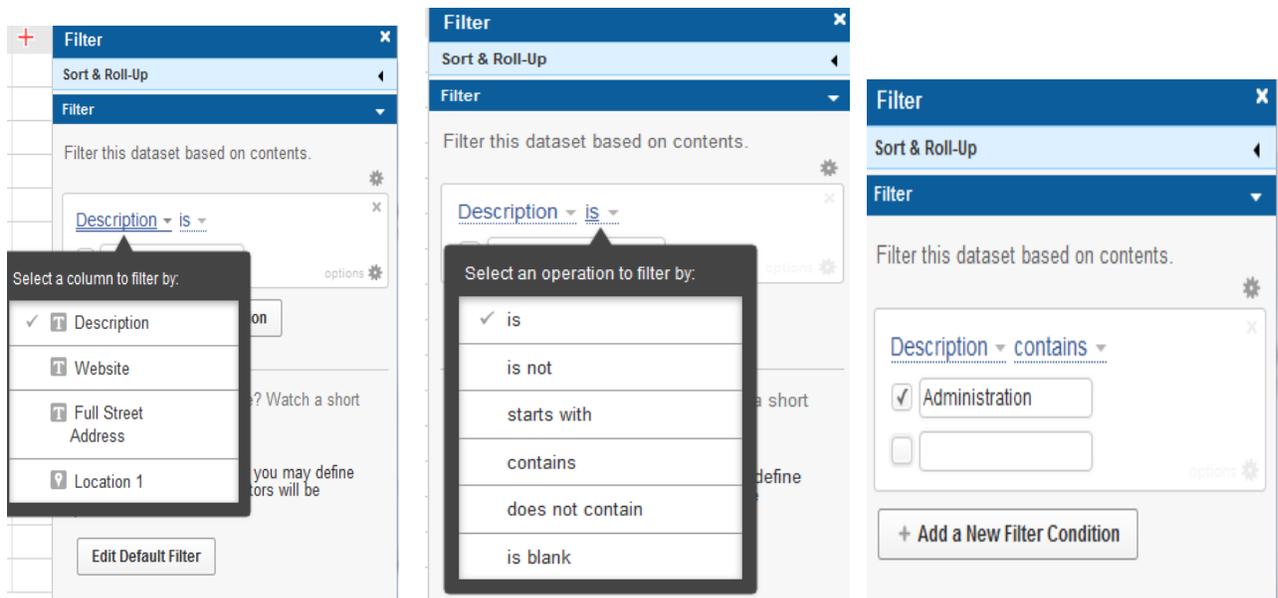
The Export Button    **Export**

Under SODA API, developers can obtain the API Endpoint which allows them to query the dataset, create new mash-ups of data, and/or create applications. Under OData, users can link to the dataset through Excel Spreadsheets. Under the Print option, users can customize the dataset print options. Under Download, users can export data in one of the listed formats.

**Export**    ✕

SODA API    ◀

OData    ◀

Print    ◀

Download    ▼

Download a copy of this dataset in a static format

| Download As | ▲ |
|---|---|
| CSV | |
| JSON | |
| PDF | |
| RDF | |
| RSS | |
| XLS | |
| XLSX | |
| XML | |

The About Button 

Under About, users can access details on the dataset through the metadata. If enabled, users are about to email the dataset owner directly with any questions.



Once completed, select the Publish Button



If you click publish dataset before being about to make updates to the dataset, select the Edit button  and Edit Dataset.  To note you must be the dataset owner or have an existing account and the dataset owner share the dataset as contributor permissions in order to edit a dataset.

Manually Importing A Dataset