Publishing a New Dataset

Purpose

This document walks through the process of creating and publishing a new dataset based on tabular or geospatial data on the Enterprise Data Platform.

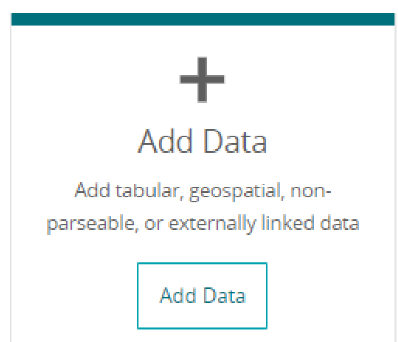
Prerequisite

Before publishing a new dataset, it is recommended that you review the Dataset Structure & Format Guide, which outlines things to consider when organizing your data and highlights the optimal file types.

Procedure

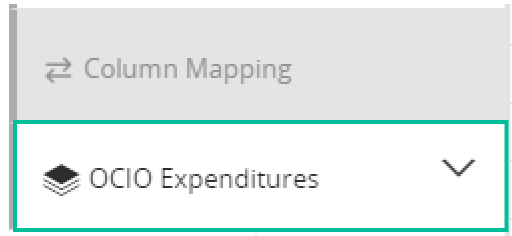
1. Log into the Enterprise Data Platform (mydata.iowa.gov) using your login credentials.
2. On the gray navigation bar at the top of the page, select “Create” then “Dataset” to begin.
3. Provide the “Dataset Name”. See Section 3.1 of the Metadata Guide for creating a title.
4. Click the “Create Dataset” button.

Selecting the Data Source



5. Under the Data Actions heading, click the “Add Data” button on the left side to select your data source.
6. Upload data file or import data from URL, both available in tabs on the right. Each method requires uploading or linking to parsable files in one of the following formats of .csv, .tsv, .xls, .xlsx, .zip (shapefile), .json (GeoJSON), .geojson, .kml, .kmz. Once your method of import has been selected follow one of the following steps:
    a. On the “Upload Data File” tab, you can drag a file onto the page or click the “Browse” button. In the new window, navigate to the file you are interested in, select the file, and click the "Open" button.
    b. On the “Import Data from URL” tab, enter the URL for the file you are interested in importing into the box, and click the “Start Import” button.

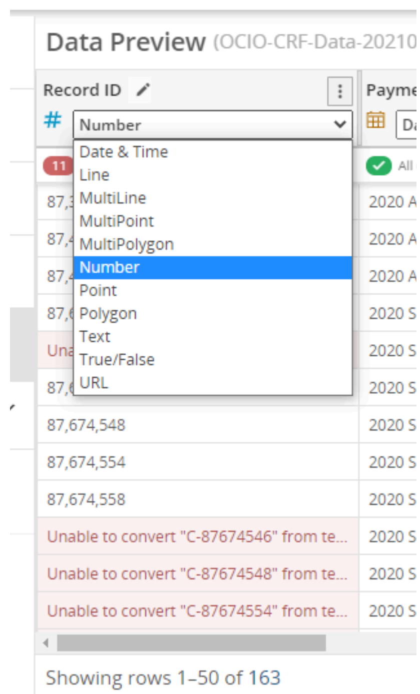
7. Once your data is imported, it will be displayed in a tabular format. If you imported a .zip, .json, .geojson, .kml or .kmz, a "geometry" field will be provided for the point, multipoint, line, multiline, polygon, or multipolygon provided in the data.



8. If your file contains multiple sheets or layers, you will now have a dropdown available in the left-hand menu bar labeled Layers/Sheets (first sheet or layer is displayed with dropdown). Click the dropdown, and choose the desired sheet or layer to import, which will refresh the data preview. *If you are interested in publishing more than one tab or layer within the file, you will need to create a new dataset for each.*

Reviewing Your Data

9. Selecting data types. The import process does its best to guess what each column represents - numbers, text, dates and times or true/false values. It does this by taking a sample from the beginning of a file and trying out every possible type for every column. The data type with the fewest errors is selected. Review the data types and override any guess by making a different selection in the data type dropdown, under the column name.

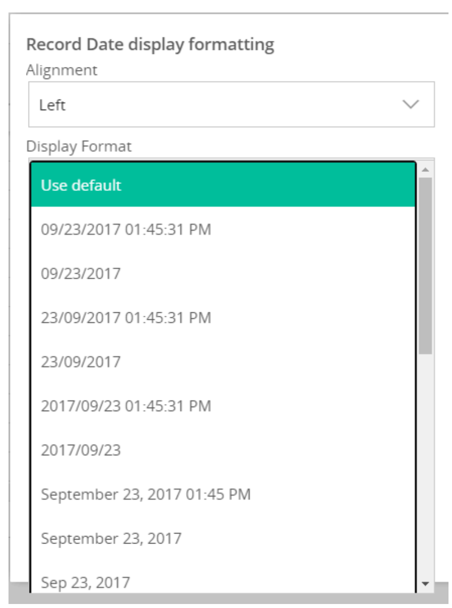


10. Review and correct errors in the data. Any records with errors will not be imported at all.
11. If the errors cannot be resolved by selecting a different data type, you may export those errors, using the “Export Errors” button in the page footer. This will only export records that could not be uploaded. The file will also contain more detail on why a row will not be imported. By exporting the errors, you may proceed with creating your dataset, fix errors externally, and append these records later by updating your dataset once it has been published. Otherwise, you can change or import a new version of the source file to this dataset by returning to step 5, and re-selecting an updated file.
12. If you widespread errors due to column mapping or character encoding, you can select the “File Configurations” menu item that allows you to provide details on the header count and position, the type of column separator used in your file, character encoding of the source file, and specify the quote character used in source file. You may also specify to trim cells' leading and trailing whitespace (recommended), and eliminate rows that are completely empty.

Adding Columns

13. If your data already contains columns with components of an address (e.g., street, city, state and zip) or latitude and longitude, you may add a georeference column to your dataset that relies on other columns already in your dataset.
    a. Click the “Add Georeference” menu option.
    b. Select source column type: lat/long, address (separated), or combined location.
    c. Use dropdown boxes to select the appropriate column in your file containing values.
    d. Specify how to handle errors.
    e. Provide display name and API field name for the newly created column. See [Section 12. Column Metadata] in Metadata Guide.
    f. Click the “Add Column” button.
14. If you would like to add a column to your dataset, you may do so by following the steps below:
    a. Click the "Add Column" menu item.
    b. Choose the Display Name, Field Name, Description, Source Column, and Data Type of the new column. See [Section 12. Column Metadata] in Metadata Guide.
    c. Click the “Add Column” button.
15. Where applicable for any new columns created, you may want to perform data transforms to enhance data. See [Section 5. Enhancing Data] in the Dataset Quality Guide for links to playbooks that give you tools and templates for transforming data. Data transforms are available as part of the column level actions that can be accessed by clicking on the vertical ellipsis icon on the appropriate column.

Configuring & Formatting

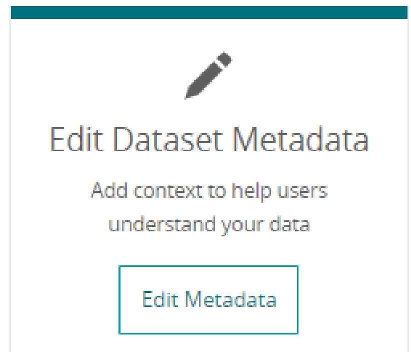


16. If you have a column that contains the record ID, click the vertical ellipsis icon on the column, then select the “Set as row identifier” option. See [Section 2.3 Record Identifier (Row Identifier)] of the Dataset Structure & Format Guide for more information.
17. If you need to hide or remove columns as needed, click the vertical ellipsis icon on the column you would like to hide or remove, then select either “Hide Column” or “Remove Column”.
18. If you need to sort your data, click the vertical ellipsis icon on the column you would like to sort by, then select either “Sort Ascending” or “Sort

Descending". Datasets may only be sorted by a single column.

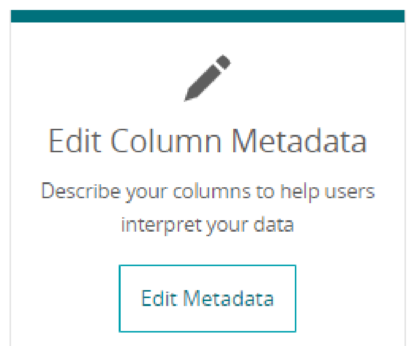
19. If you need to format your data, click the vertical ellipsis icon on the column you would like to format, then select "Column Formatting". With column formatting you may specify how you would like the data aligned (i.e., left, center, right), and how you would like the data displayed. The available display formats will be dependent on the data type of the column.
20. Once you have finished configuring and formatting your dataset, click the "Save Draft" button in the bottom right corner of the page.

Adding Dataset Metadata



21. Under the Data Actions heading, on the "Edit Dataset Metadata" block, click the "Edit Metadata" button.
22. Enter appropriate information into the fields on the provided forms. The [Metadata Guide](#) provides information and instructions on what should be included in each field.
23. Once the form is complete, click the "Save" button at the bottom of the page.

Adding Column Metadata



24. Under the Data Actions heading, on the "Edit Column Metadata" block, click the "Edit Metadata" button.
25. Rename column names where needed and provide column descriptions. See Section 12. Column Metadata in Metadata Guide.
26. Reorder columns if needed by hovering over the icon under position for the row that describes the column, and left click it to grab it. Drag and drop the column where you want it. The column presented at the top of this page will be the left most column in the dataset. The column presented at the bottom of this page will be the right most column in the dataset. See Section 5. Order of the Dataset Structure & Format Guide.
27. Once complete, click the "Save" button at the bottom of the page.

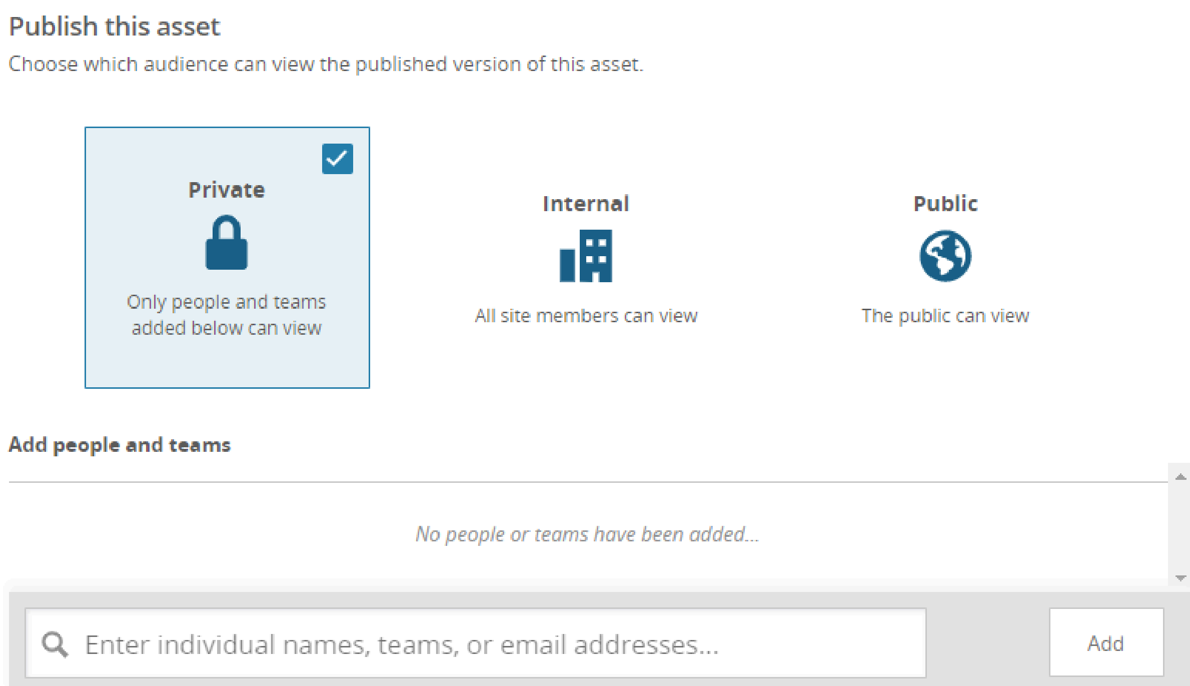
Adding Collaborators

28. If you want others to review the dataset, you may add collaborators by clicking the "Manage Collaborators" button in the upper right hand corner of the page. If not, proceed to step 31.

29. Enter individual names, teams, or email addresses for those you would like to have access to the draft dataset, and click the “Add” button.
30. Once the review by collaborators is complete, proceed to the next step.

Publishing the Dataset

31. Click the “Publish Dataset…” button in the upper right corner.
32. On the “Review Changes” pop-up, click the “Continue” button on the lower right corner.
33. On the “Publish this asset” pop-up, choose which audience can view the published version of your dataset by selecting private, internal or public. [See Section 13. Granting Access](#) to Assets of Data Portal Governance Guide.



34. If you intend to maintain the dataset as private, you may give others access to it by entering individual names, teams, or email addresses in the text box, then click the “Add” button.
35. Click the “Publish” button in the lower right corner.
36. Once your dataset is finished processing, you may navigate to the primer page, by clicking the “Go to Primer” button on the pop-up.