ClairCity data portal
http://claircitydata.cbs.nl

User Guide

Version 20180419, replaces version 20161206

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 689289.

Contents

Introduction 3

The data portal in a nutshell 3

Searching for data 5

Using the ClairCity model 5

Using the Search engine 6

Using the list of organizations 7

Using the list of groups 8

Uploading a dataset 9

Assigning your dataset to the right Group(s) 11

Metadata guidelines 12

Title and URL 12

Description 13

Tags 14

Organization 14

Visibility 15

Quality aspects 15

City 15

License type 15

Groups 15

The Application Programming Interface (API) 17

How to register 17

Contact 17

Introduction

The goal of the ClairCity data portal is to organize data produced and consumed by the ClairCity organizations within the ClairCity project. This document is a brief guide for getting started. Additional guides explaining more advanced topics may be added later.

The data portal in a nutshell

The ClairCity data portal organizes data in so called *Datasets*. Each dataset has some standard metadata fields such as a *title*, *description*, *tags*, *visibility*, *quality aspects* and *license*. Each dataset is owned by exactly one of the 16 ClairCity organizations or by the ClairCity organization as a whole.

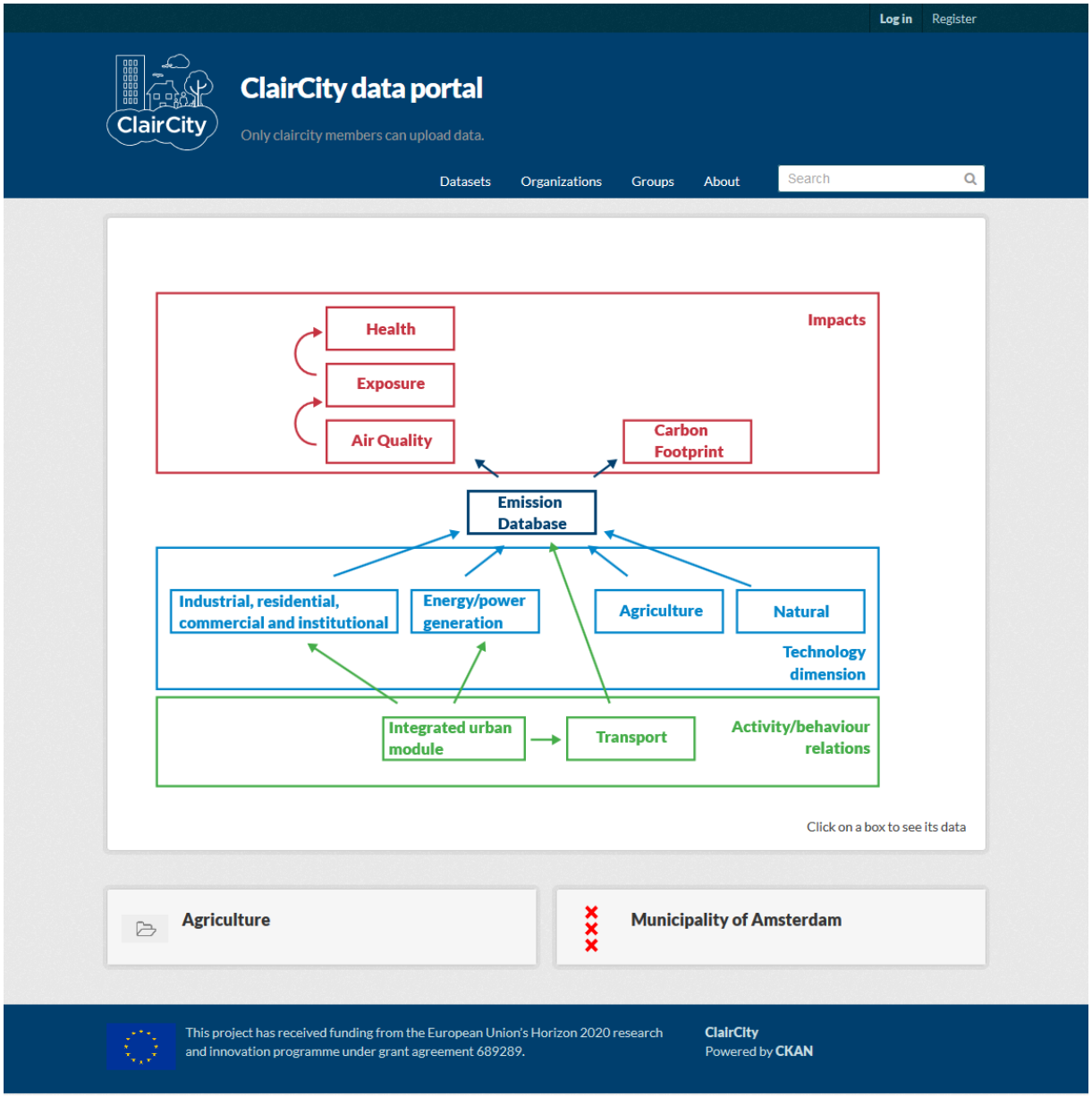
Datasets in the ClairCity data portal have some ClairCity specific metadata fields. Each dataset may be added to one or more *Groups* from the ClairCity *conceptual model* (see next page), such as Agriculture, Transport or Health and may be assigned to one of the ClairCity *Cities* (Liguria, Aveiro, Ljubljana, Amsterdam, Bristol).

Only members of the ClairCity project are allowed to add datasets. They can effectively choose from three levels of access control: *public*, *private* to the Claircity organization and *private* to the owner organization. More details about this mechanism can be found in the section “Uploading a dataset”.

The data portal offers a data preview in a datagrid, chart or map. This works if the data was recognized (csv, json or xml) and contains the right type of variables for the preview type (like lat/lon coordinates or a geojson field for maps). Other preview facilities may be added later. If the portal cannot recognize or parse the data, it is stored as a blob for download. We advise to use one of recognizable formats as much as possible. If you have ideas or suggestions for new ClairCity specific previews for your data, please contact us.

It is not allowed to add any privacy sensitive data to the data portal. It is the responsibility of the user to verify that a dataset does not contain any privacy sensitive data and that it does not infringe intellectual property rights.

Below is a screenshot of the front page of the data portal as it is displayed for non registered users. The main menu (Datasets, Organizations, Groups, About) and the search field is always present. Right under the main menu there is the ClairCity conceptual model. Next to the conceptual model a selection of data groups and / or organisations is displayed depending on the public data available in the portal.

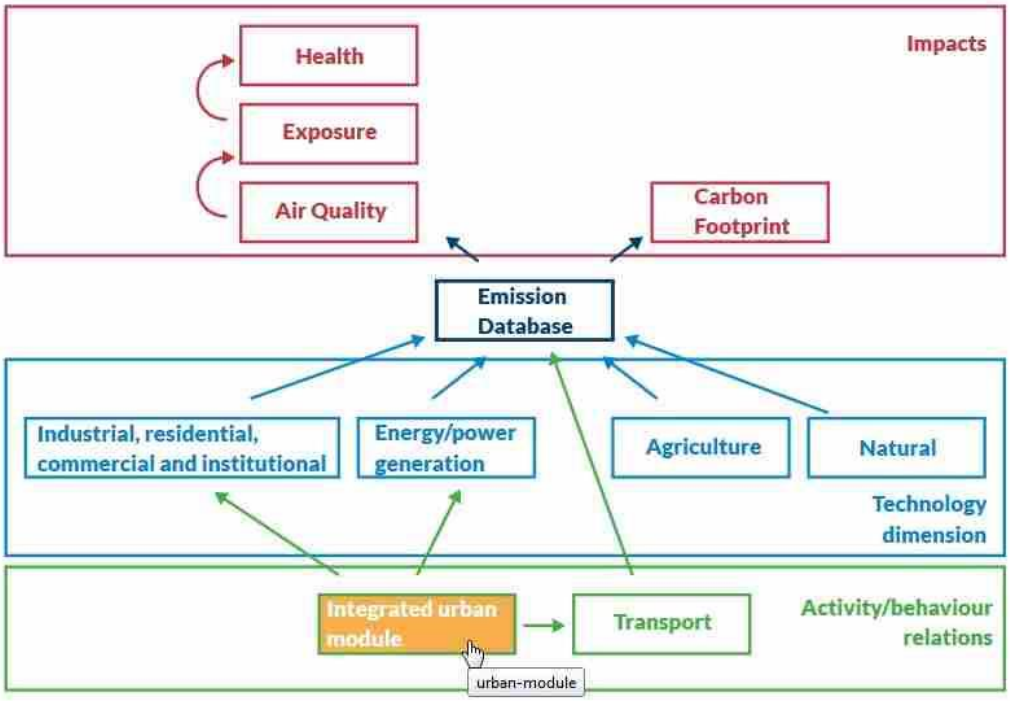

Log in
Register
ClairCity
ClairCity data portal
Only claircity members can upload data.
Datasets
Organizations
Groups
About
Search
Impacts
Health
Exposure
Air Quality
Carbon Footprint
Emission Database
Industrial, residential, commercial and institutional
Energy/power generation
Agriculture
Natural
Technology dimension
Integrated urban module
Transport
Activity/behaviour relations
Click on a box to see its data
Agriculture
Municipality of Amsterdam
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 689289.
ClairCity
Powered by CKAN

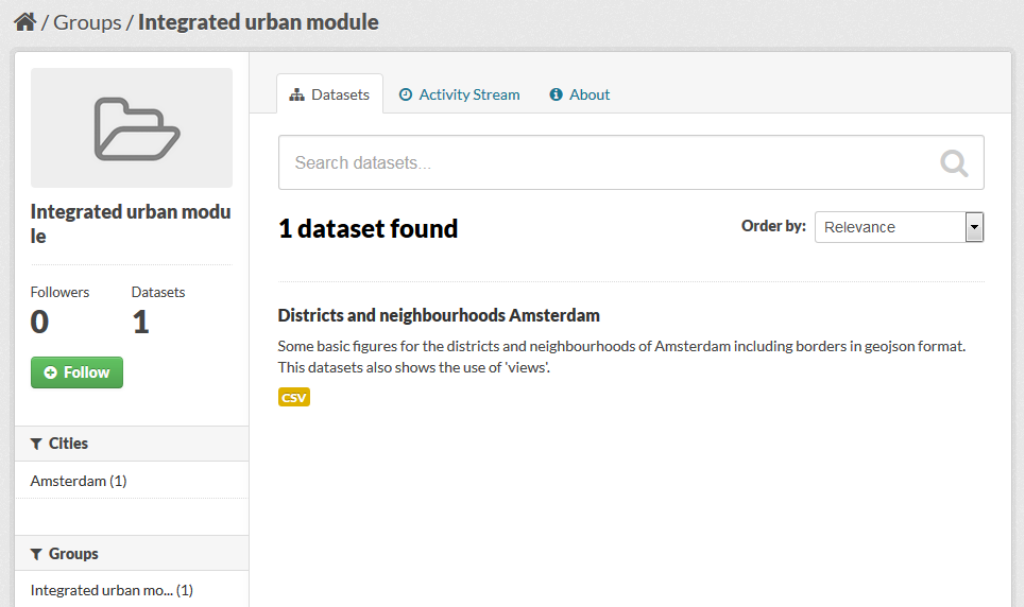
Searching for data

The data portal offers various mechanisms to search for data:

Using the ClairCity model

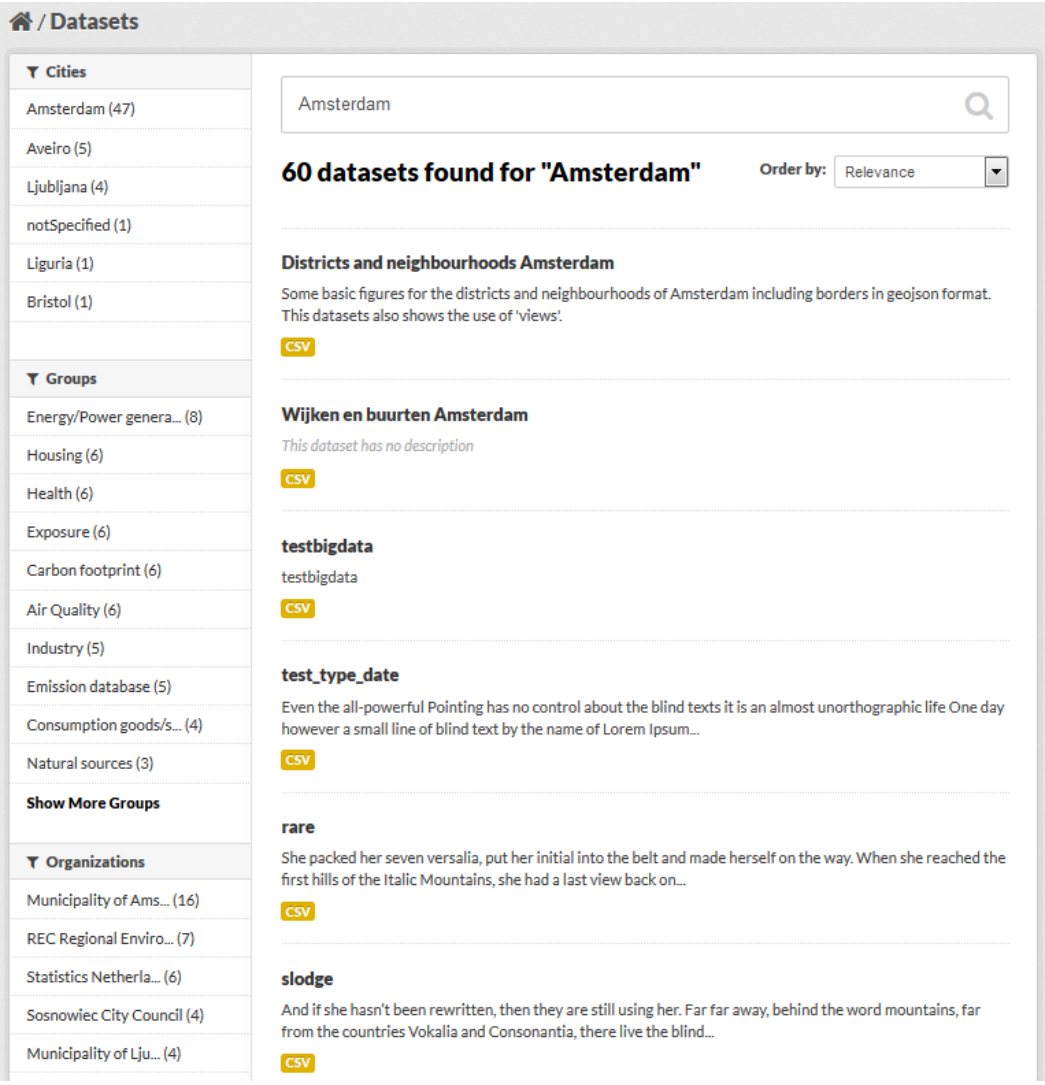
In the conceptuel model from the ClairCity project, displayed on the home page of the ClairCity data portal each box represents a group of datasets. A click on a box brings the user directly on a page displaying all datasets of that group visible for that user. The screenshots below the selection of the datatypes for the “Integrated Urban module”:





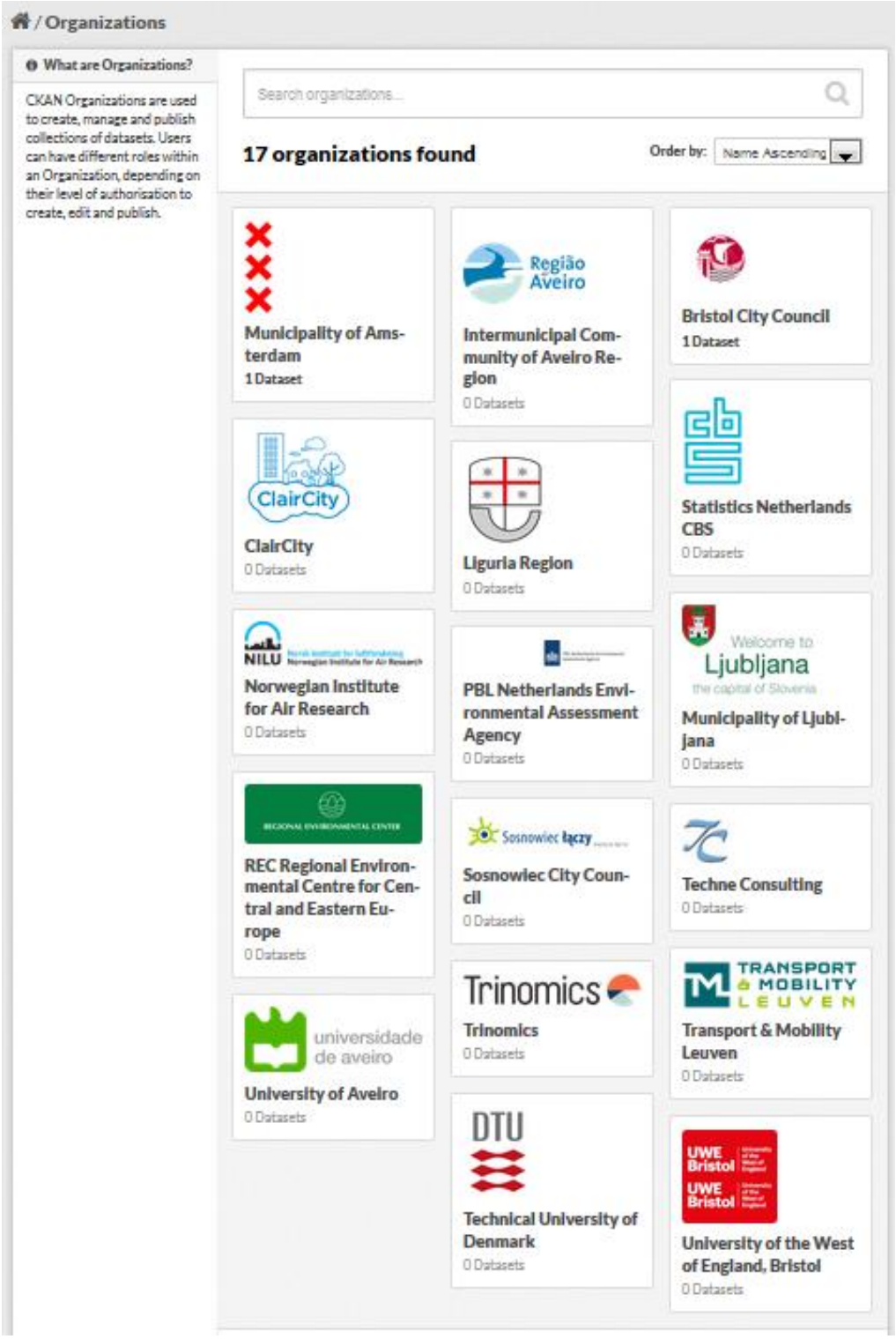
Using the Search engine

The ClairCity data portal contains a seach engine. Using the search box or clicking the “Datasets” tab gives a search results page with facets for cities, groups, organizations, tags and formats. The screenshot below shows such search results page for the search phrase “Amsterdam”:



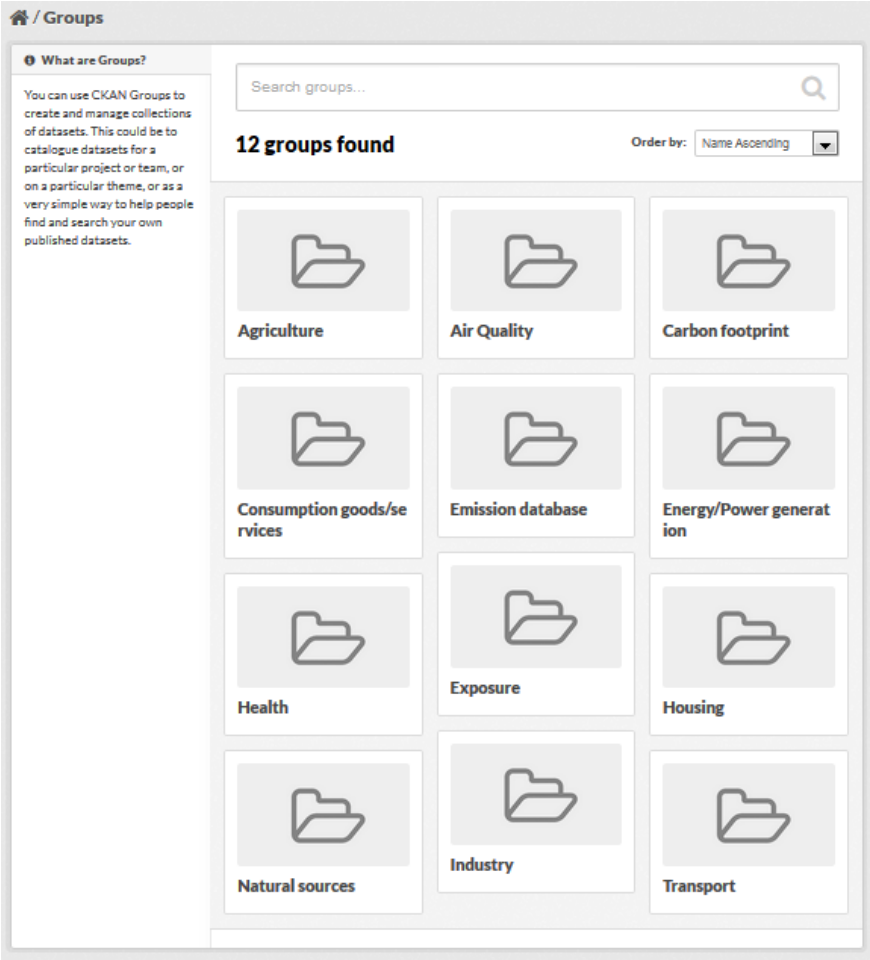
Using the list of organizations

The “Organizations” tab brings you to the page which lists all ClairCity organizations and their respective number of datasets. A click on an organization brings you to all datasets owned by that organization. Here is a screenshot:



Using the list of groups

The “Groups” tab brings you to the page which lists all groups of datasets from the conceptual model. A click on an group brings you to all datasets for that group. This is the same functionality as clicking on a group box on the conceptual model on the home page, which might be a faster and more intuitive way of selecting a group. For completeness, here is a screendump:



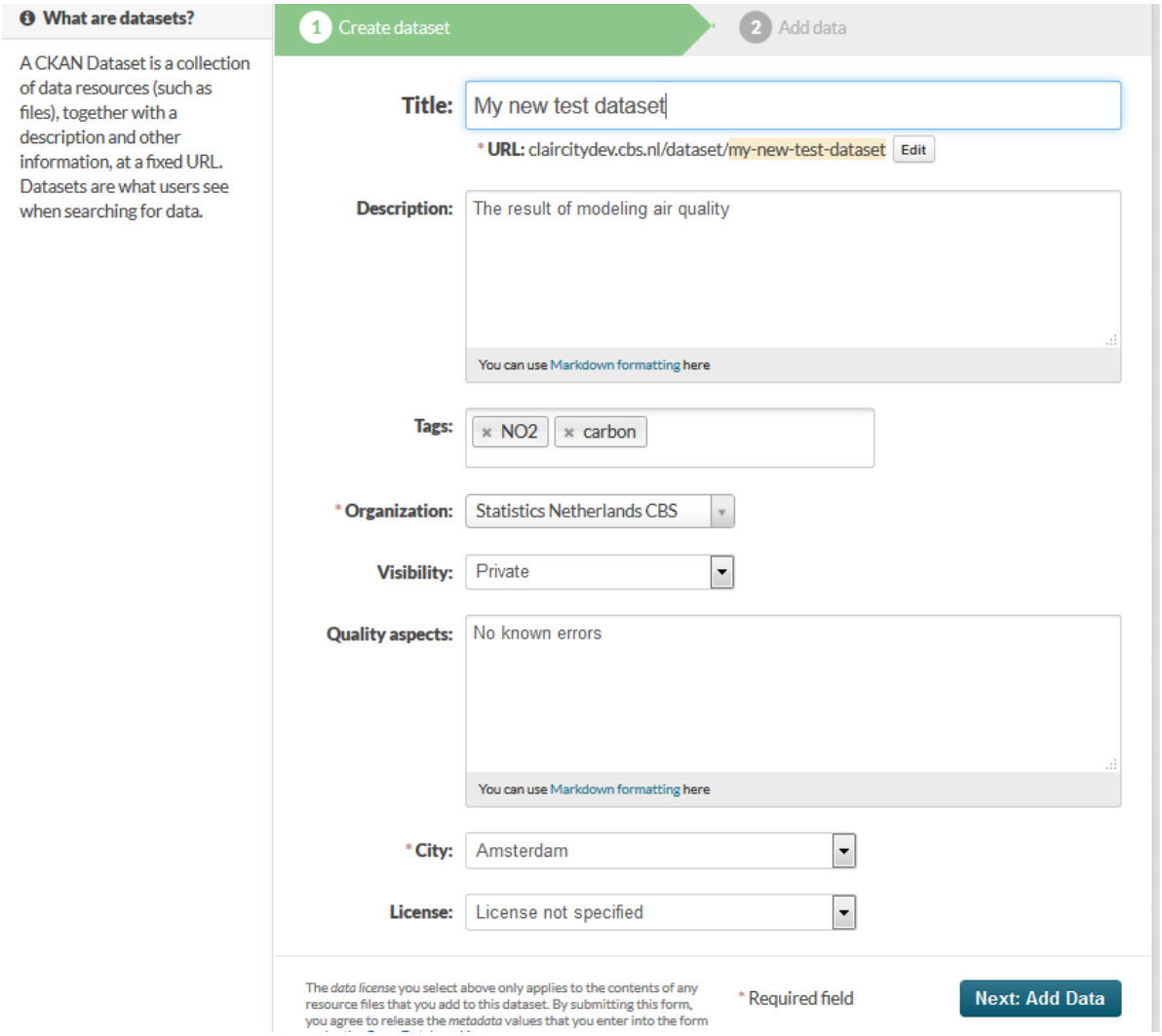
Uploading a dataset

Registered users assigned to one of the ClairCity organizations have an “Add Dataset” button on the datasets page. This brings you to the “Create dataset” form where you specify the metadata of your dataset. The fields are:

- **Title**
- **Description**: short descriptive text
- **Tags**: you can create your own
- **Organization**: either your own organization or the ClairCity organization.
- **Visibility**: Public or Private. Public makes your dataset visible to the world, Private will make your dataset visible to the organization you specified above.
- **Quality aspects**
- **City**: the city the data applies to (use not specified if it is not specific to one city)
- **License**

Note: Choosing the right combination of the Organization and Visibility fields you can choose between three levels of access control as specified in this table:

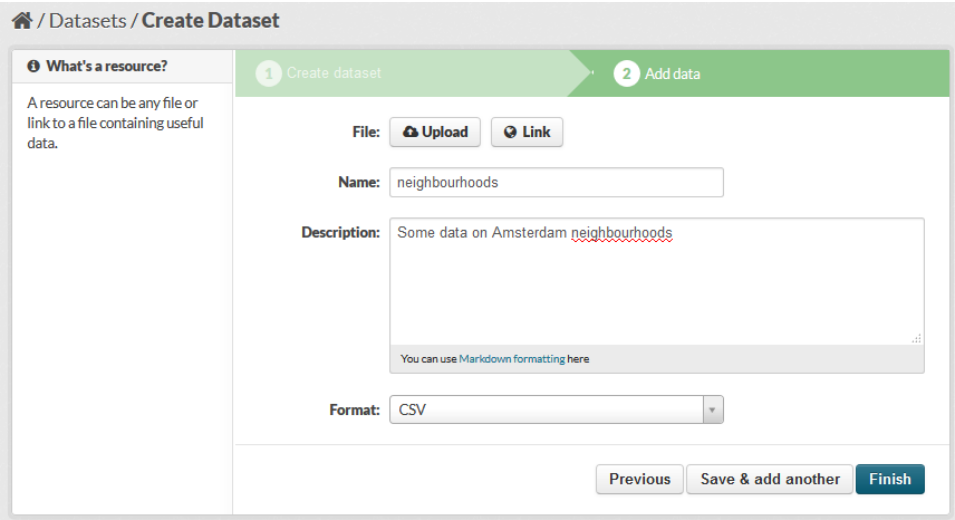
Organization	**Visibility**	
<your organization>	Public	Open to the world
<your organization>	Private	Visible to the members or your organization
ClairCity	Public	Open to the world
ClairCity	Private	Visible to all members of ClairCity



After creating your dataset, the portal invites you to add data resources to this dataset in the "Add data" form. The fields are:

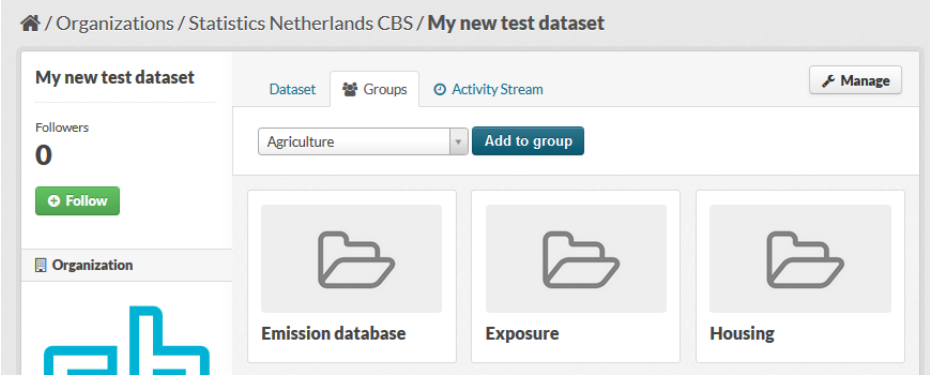
- **File**: either an upload or a link, we advise not to upload any data if you can link to it.
- **Name**
- **Description**
- **Format**: CSV, JSON or XML. If empty the system tries to recognize it. Leave empty in case of another format

A dataset can have multiple data resources, which might be useful to group several files on the same subject together. One example is to provide the data in a CSV as well as a textual description into one dataset Another example is to provide the data itself as well as a link to where the data was derived from into one dataset.



Assigning your dataset to the right Group(s)

The “Create dataset” form does not offer you a way to assign it to one or more groups immediately. So after creation it is not assigned to any of the groups. You can however assign your dataset to the right Group(s) afterwards using the “Groups” tab of the dataset page as displayed here:



Metadata guidelines

The origin of data resources on the internet may often be obscure. For data resources in research and for archiving data in general, quality criteria are described in the FAIR principles. In the FAIR Data approach, data should be:

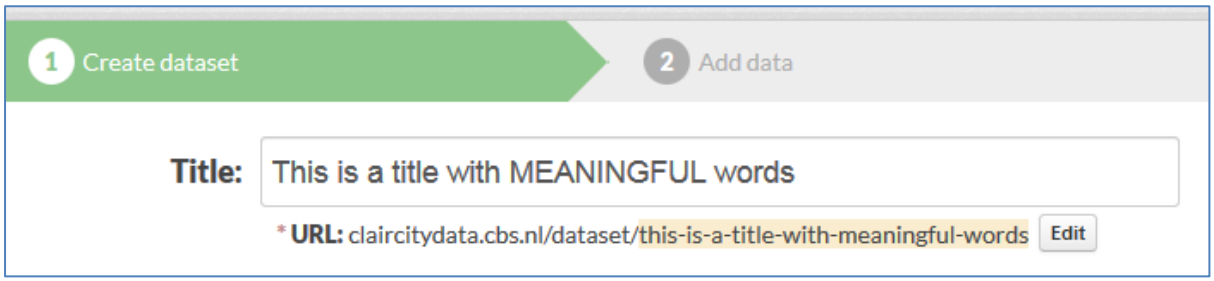
- **F**indable – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets
- **A**ccessible – Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content
- **I**nteroperable – Ready to be combined with other datasets by humans as well as computer systems
- **R**eusable – Ready to be used for future research and to be processed further using computational methods

Since ClairCity aims to develop a new model for ir quality improvements throughout Europe, it would be desirable to meet these FAIR criteria to a high degree.

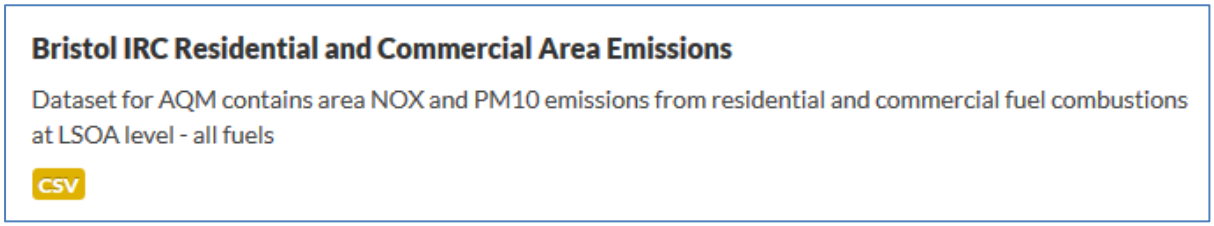
We propose to all parties that are processing and uploading data to the ClairCity data portal to take into consideration the guidelines per metadata field described below.

Title and URL

Use a title that helps search engines and people to understand what the dataset is about. Keep in mind that the title is used to generate a unique URL to your dataset. See the picture below. It is however possible using the **edit** button to overrule the generated URL. We advise not to do this and use a title with meaningful words.

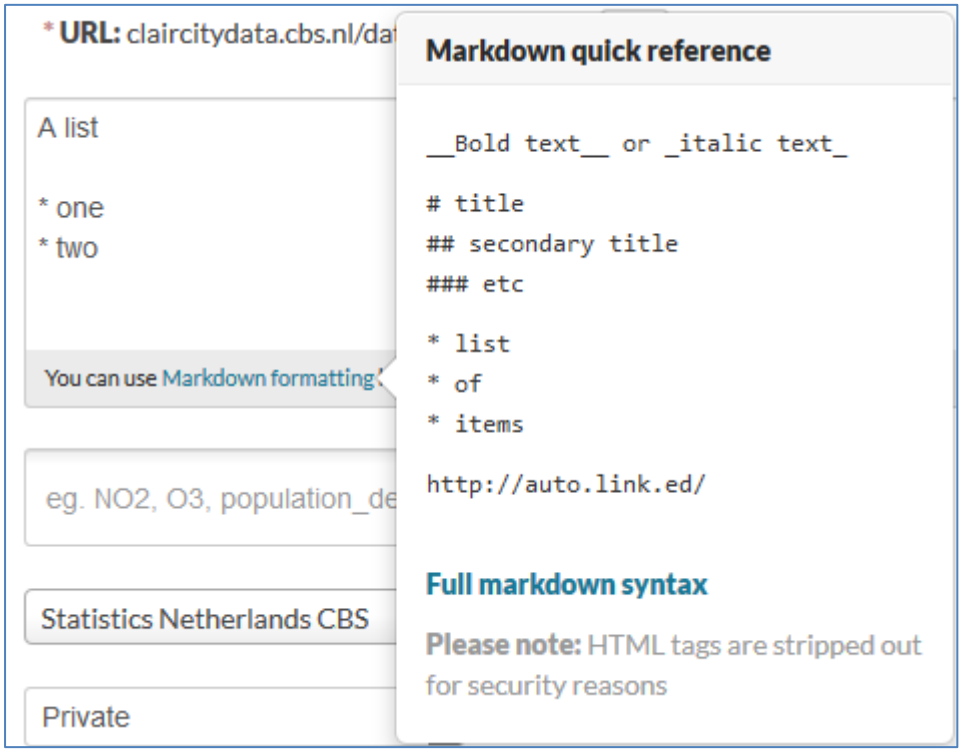


The screenshot below shows an example of a good title.



Description

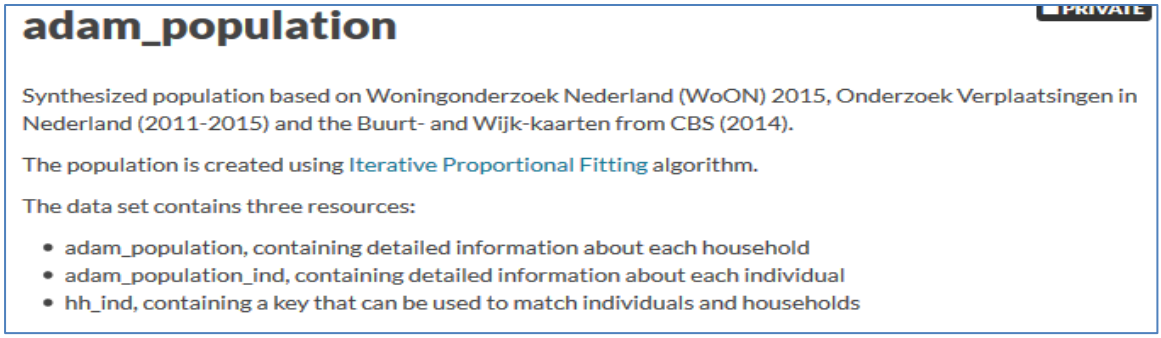
This is a free description field where you can use Markdown, which contains some formatting options for **bold**, *italic*, Headers, bullets, and hyperlinks. The screenshot below shows the Markdown possibilities in this field.



We suggest to always include a description with at least information on:

- Origin of the data
- Time period data applies to
- Collection method
- Sampling design (in case of surveys)
- Analytical methods / algorithms
- Unit of measure

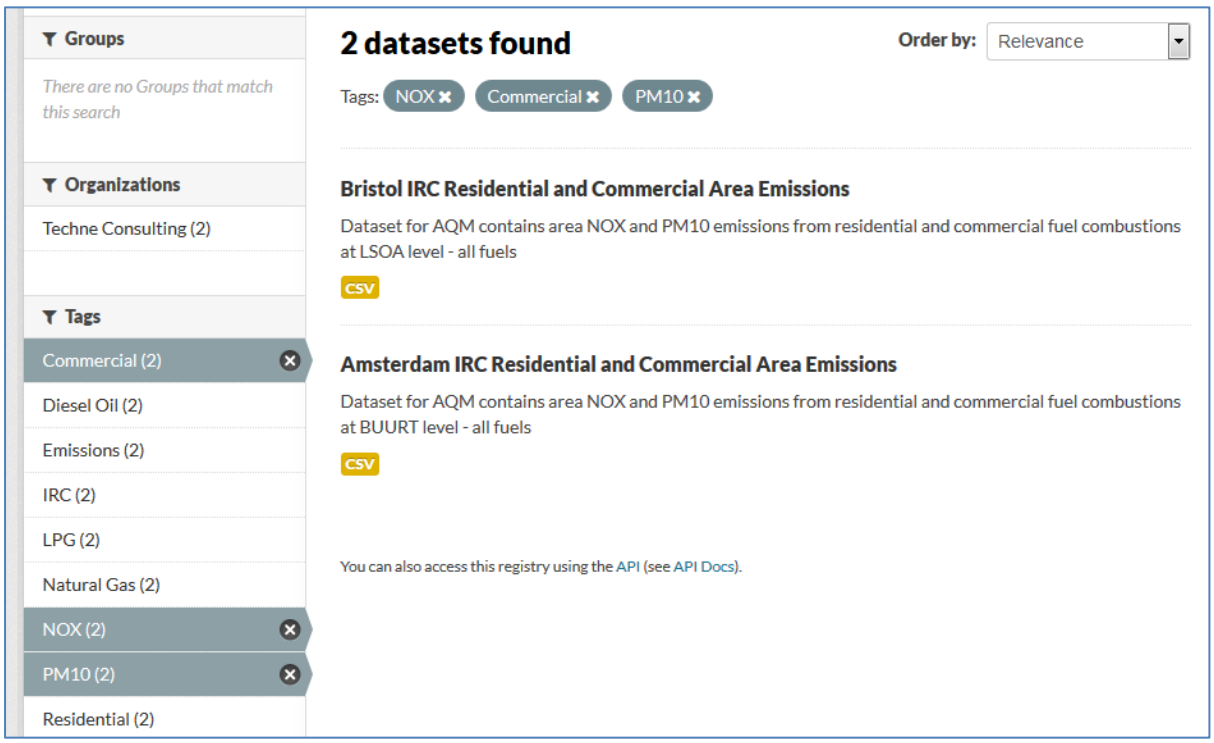
The screenshot below gives an example with a description of the data sources, the time periods, and a hyperlink to the algorithm being used.



There is also a Description field on the level of resources in a dataset. This can be used for resource-specific metadata. We suggest adding all general metadata for a dataset as much as possible in the dataset Description field.

Tags

Tags can be defined freely by the data uploaders. You can reuse already defined tags by others or define tags yourself. Reusing tags enables all of the FAIR principles for data management. The ClairCity data portal search engine can be used for tag-based search as is displayed below. Here you see two datasets share the tags NOX, Commercial and PM10. Other datasets use LPG, IRC, and Natural gas.



We suggest that you search for appropriate existing tags for your dataset first, before defining your own. However creating your own tags is a good approach if you add a dataset which is incomparable to the ones that are already available.

Organization

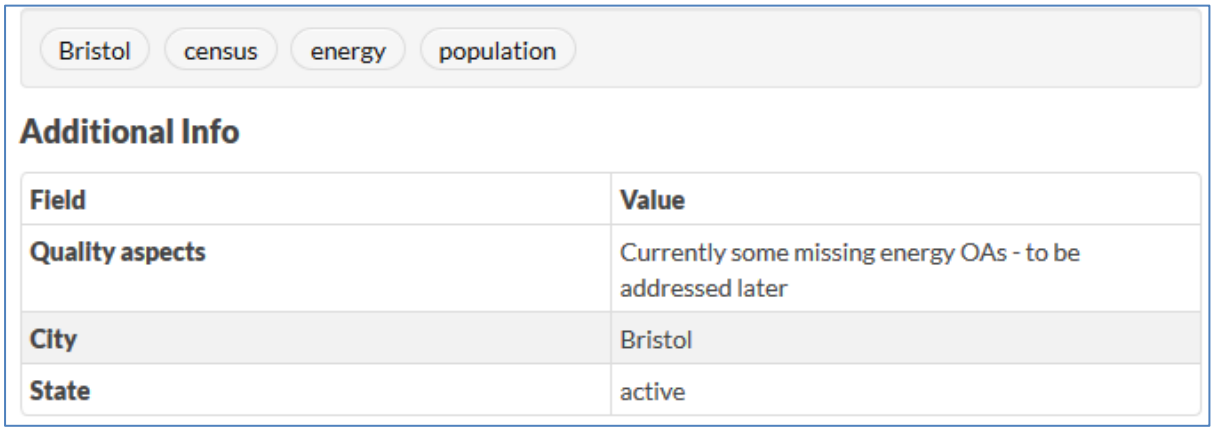
Use your own organization or the ClairCity organization in case you want to share with the ClairCity partners only. See the table on combination of organization and visibility in the chapter ‘Uploading a dataset’.

Visibility

To support the applicability of the FAIR principles we suggest making your dataset public if possible. Only public datasets can be indexed by internet search engines.

Quality aspects

This is a free description field where you can use Markdown. This field may be used to enter quality considerations not naturally fitting in the dataset field Description, but important in regard to knowledge for others interpreting your data. The screenshot below shows a good example.



City

Use one of the 6 ClairCity cities if your dataset applies to one specific city. For general data like emission factors per vehicle, usable in any city, leave it ‘not specified’.

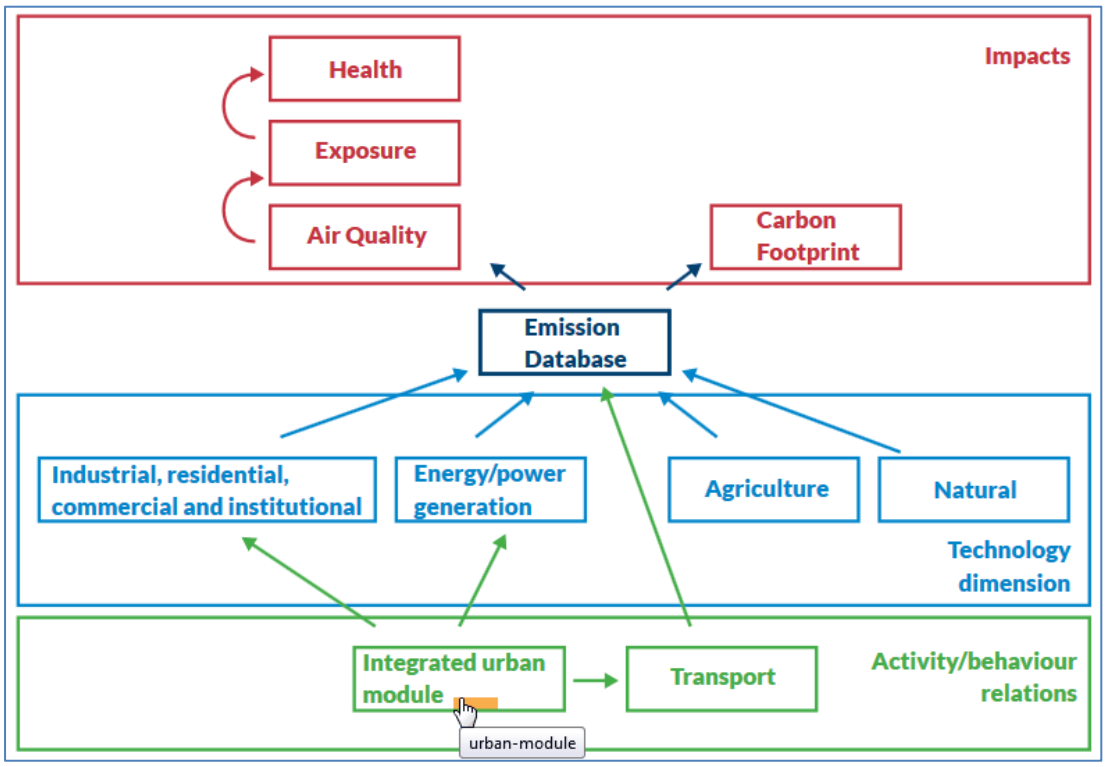
License type

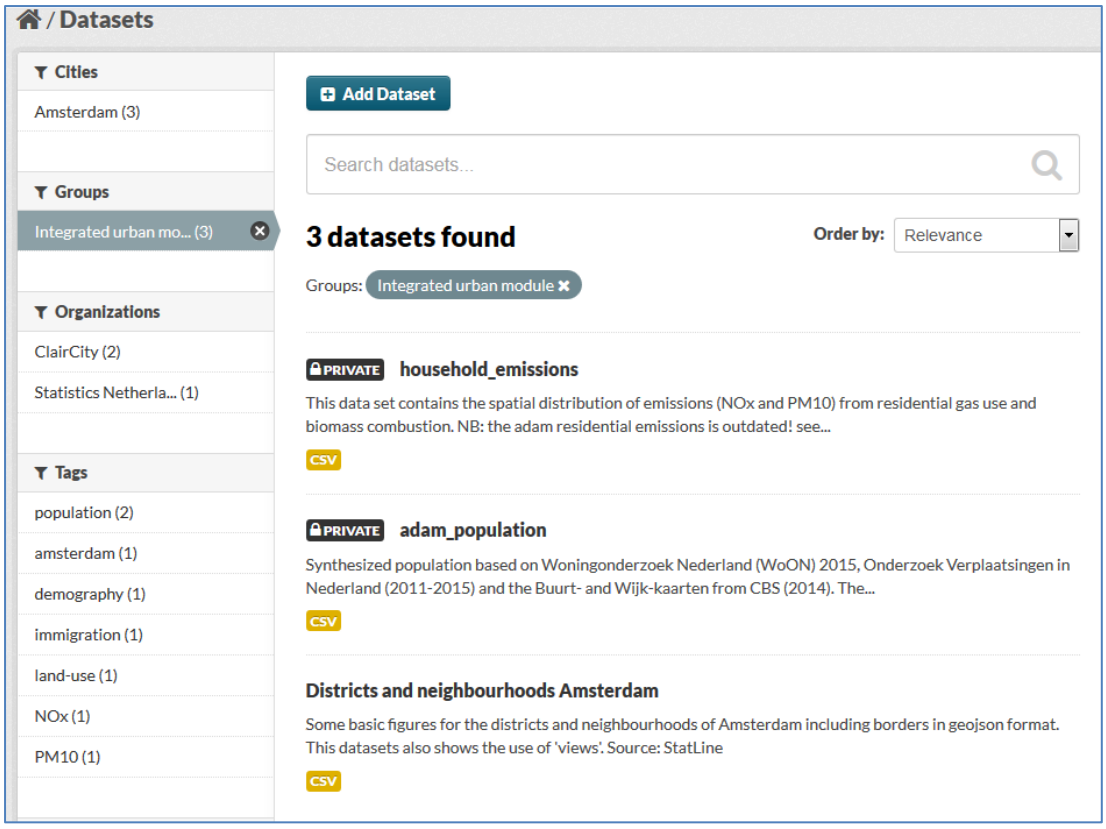
For your dataset to truly be useable according to the FAIR principle it is very recommendable to choose an open source license so that others know how they can reuse your data. For choosing a license you might for instance consult https://choosealicense.com. In the data portal there is a list of preconfigured common licenses available. If your license is not there, you can choose ‘Other’ and add your license file as a resource to your dataset.

Groups

The data portal supports the assignment of datasets to one or more of the groups / boxes in the conceptual ClairCity model displayed on the front page. This is described in the chapter on ‘Assigning your dataset to the right Group(s)’. Assigning your dataset to a group makes it

findable with one click via the conceptual model diagram on the homepage. See the screenshots below.





The Application Programming Interface (API)

The data portal offers programmatic access to a resource of a recognized format via an application programming interface (API). This so-called DataStore API can be used for reading, searching and filtering data without the need to download the entire file first. One can use this mechanism to execute queries against data stored in the data portal. To do so one needs an api key, which by default has any registered user or can can be regenerated from your personal settings page. More information on the possibilities of the DataStore API can be found here:

http://docs.ckan.org/en/ckan-2.6.0/maintaining/datastore.html#the-datastore-api

Contact us if you want to interact with the ClairCity data using this API. We have example software to do this and can together decide on the optimal strategy for your use case.

How to register

Click on the register button and follow the wizard. Once you created an account please send an email to claircity@cbs.nl mentioning your username and the ClairCity organization you work for. We will assign your account to this organization in the data portal and to the ClairCity organization. From then you have the right to upload datasets on the ClairCity data portal on behalf of your organization.

Contact

Questions, suggestions, remarks can be sent to claircity@cbs.nl