

510 Data Catalog Manual

Version 0.200.0



An initiative of
the Netherlands
Red Cross

510 Data Catalog

Table of Contents

[How to Log in](#)4

[Adding or editing a dataset](#)6

[Finding a dataset](#)23

[Frequently asked Questions](#)28

510 Data Catalog

Introduction

Welcome to the 510 Data Catalog Manual! Here you will be able to learn about how to use the data catalog i.e. to add datasets in the catalog and to find data in the catalog.

The 510 Data Catalog provides an overview of all (important) datasets in use within 510. It enables the 510 team members to share information about the data they are using with each other and other stakeholders.

You can add the data after logging in. When you search for data in the data catalog you can see information about the data and can also see where it sits in the database or the datalake.

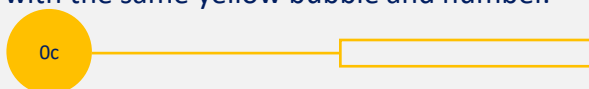
Intended users: 510 staff, consultants, volunteers (anyone making use of technical architecture, making pipelines or storing data in the 510 cloud).

Features of the Application:

- The user can access certain data without logging in
- The user can access all the data after logging in
- The user can add data to the catalog after logging in.

Legend:

Yellow circles with a line and box will indicate different features, which will be explained in text with the same yellow bubble and number.



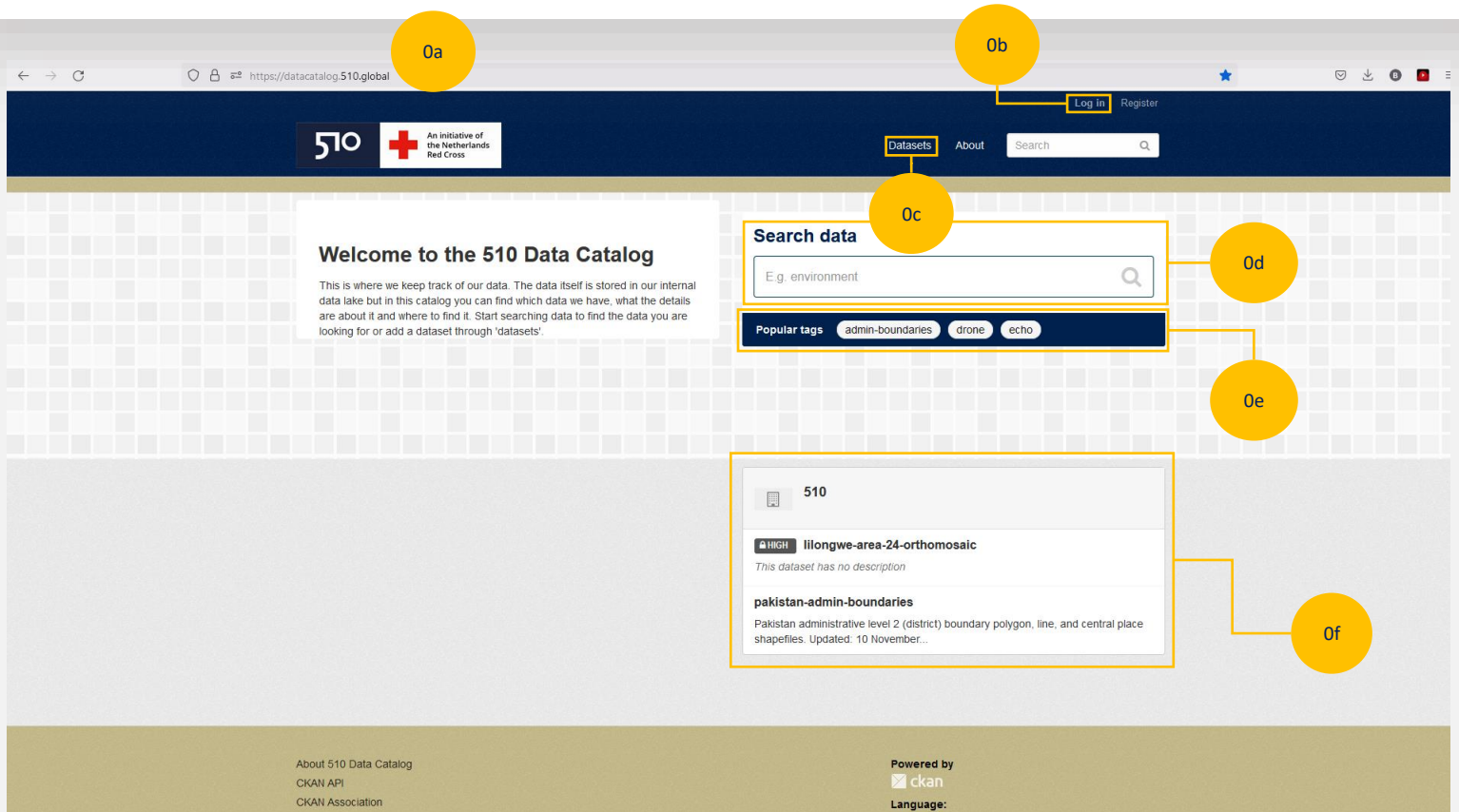
WHAT DOES THIS BOX MEAN?
Explanation of actions.

** This is the manual for the application as per the last visible update on 01-07-2022*

510 Data Catalog

How to use the 510 Data Catalog

Home screen



0. Home:

- a. The URL of the Data Catalog is <https://datacatalog.510.global>
- b. Click to log in to your account.
- c. Click to access datasets.
- d. Type in text about the dataset you are looking for and click enter to search.
- e. The most common tags for the datasets can be found here.
- f. Shows you recent datasets.

WHEN DO YOU USE THIS SCREEN

You use this screen to login to 510 Data Catalog or look for a dataset.

510 Data Catalog

How to use the 510 Data Catalog

Log in screen

https://datacatalog.510.global/user/login

Log in Register

510 An initiative of the Netherlands Red Cross

Datasets About Search

Home / Login

Need an Account?

Then sign right up, it only takes a minute.

Create an Account

Forgotten your password?

No problem, use our password recovery form to reset it.

Forgot your password?

Login

Username:

Password:

☒ Remember me

Login Login via Red Cross AD Account

About 510 Data Catalog
CKAN API
CKAN Association
OPEN DATA

Powered by ckan

Language:
English (United Kingdom)

i.

0b. Log in:

i. Click here to log in using the Red Cross AD Account.

WHEN DO YOU USE THIS SCREEN

You use this screen to login to 510 Data Catalog.

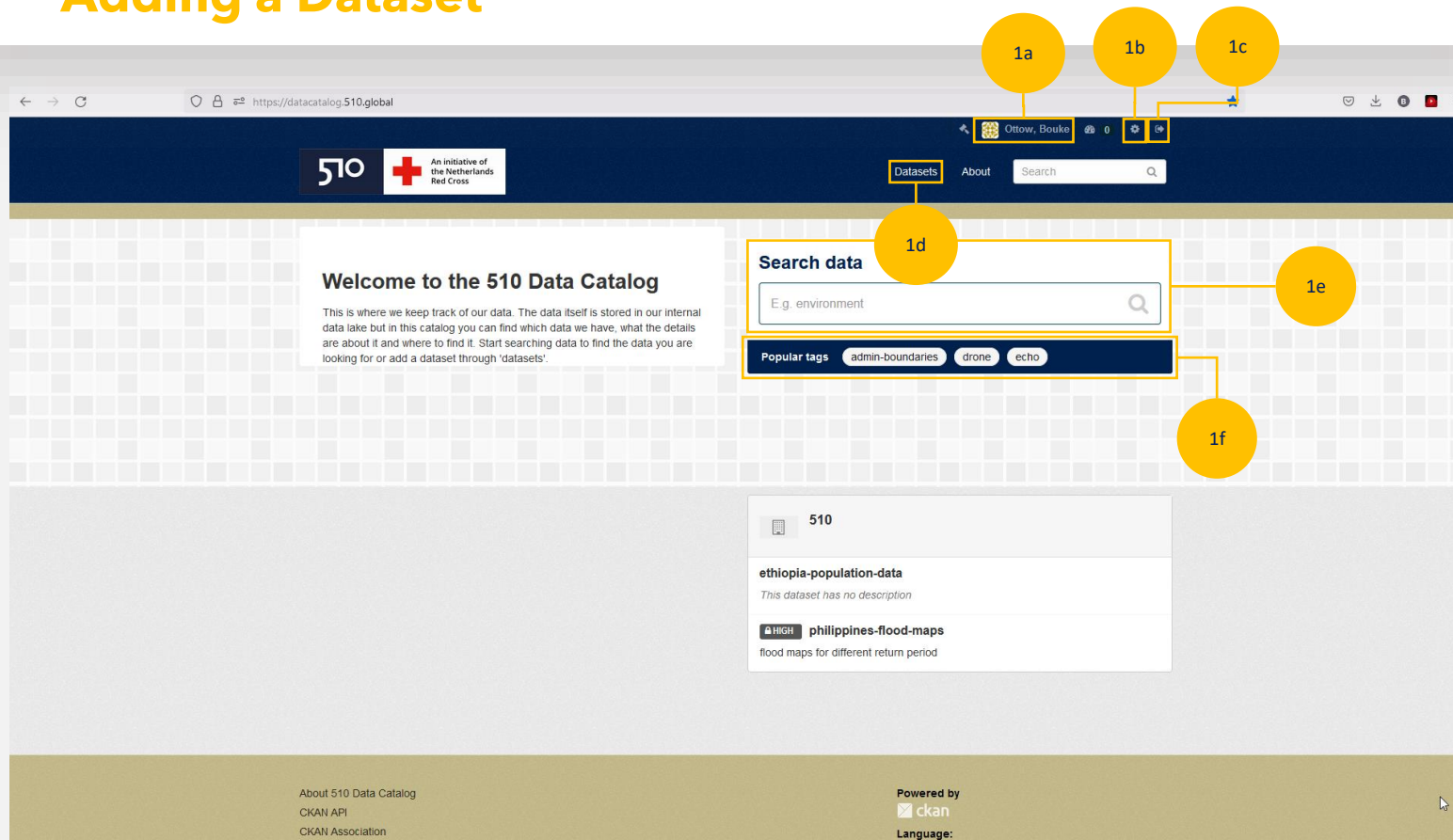
510 Data Catalog

Adding or editing a dataset

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset



1. Home:

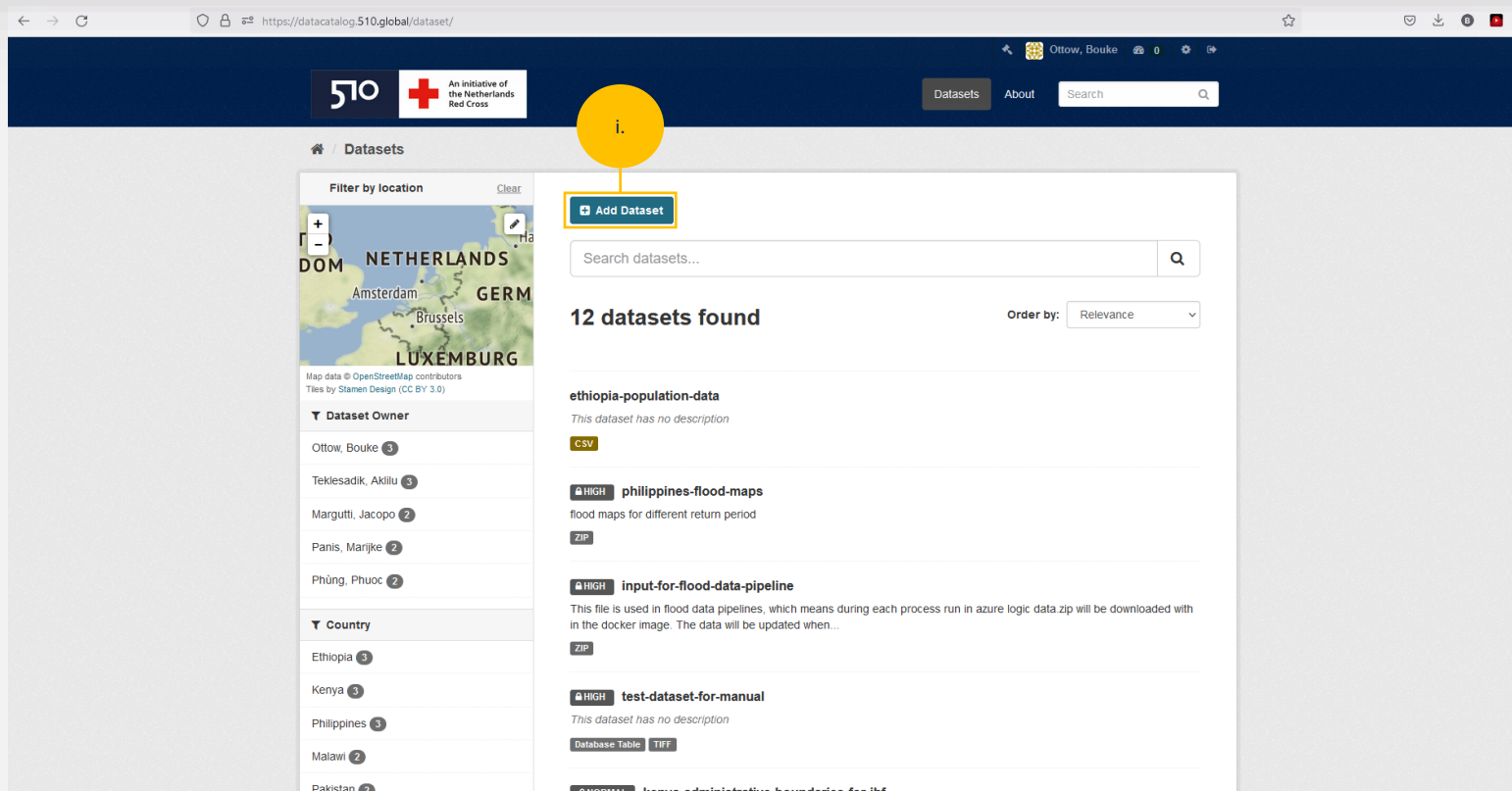
- a. Click here to see information about your account and datasets.
- b. Click here to change account details.
- c. Click here to log out.
- d. Click here for an overview of all the datasets.
- e. Type here to directly search for a dataset.
- f. Click a tag to see all datasets with that tag.

WHEN DO YOU USE THIS SCREEN
You use this when you go the home screen after logging in.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset



1d. Datasets:

i. Click here to add a new dataset.

WHEN DO YOU USE THIS SCREEN

You use this when going to the datasets page to add a new dataset.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

The screenshot shows the 'Create Dataset' form in the 510 Data Catalog. The form is divided into two main sections: 'Create Dataset' (left) and 'Dataset name' (right). The 'Create Dataset' section contains instructions on how to create a dataset and a list of fields to fill in. The 'Dataset name' section contains fields for 'Dataset name', 'URL', 'Description', 'Tags', 'License', 'Organisation', and 'Source'. Annotations are present: '2a' points to the 'Create Dataset' section, '2b' points to the 'Dataset name' section, and '2c' points to the 'Description' field. A yellow arrow points from the 'Description' field to the 'Dataset name' field.

2a

2b

2c

Create Dataset

There are two important entities in this catalog.

1. the dataset
2. the resource

The dataset includes all important (background) information about the dataset. The resource has information (location, amount, type, etc.) about the specific files related to it. A resource is always part of a dataset and one dataset can have multiple resources.

Fill in all fields to add a new dataset to the 510 data catalog. Additional information about the fields can be found below.

Dataset name

For the Dataset name you should choose something that is understandable for people searching for data and is short and concise. There are no rules on what needs to be in it. Just remember that many of the details will be in the other metadata fields so they don't all need to be included in the name.

The dataset name is stored as lower case and a whitespace is replaced by a dash '-'.

Description

1 Create dataset **2** Add data

Dataset name:

eg. data 510

* URL: datacatalog.510.global/dataset/<dataset> Edit

Description:

eg. Some useful notes about the 510 data

You can use Markdown formatting here

Tags:

eg. economy, mental health, government

License:

License not specified

License definitions and additional information can be found at <http://opendefinition.org/>

* **Organisation:**

510

Source:

2. Create Dataset:

- a. Read this to understand how a dataset is built up.
- b. Fill in all the metadata fields to describe the dataset and...
- c. look up information about the field to understand what is expected of you.

WHEN DO YOU USE THIS SCREEN
You use this when creating a new dataset.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

Always select just 510, as is the default. So no need to change this.

Source

Put here the source of the dataset. If it is downloaded from a website you can put the url, otherwise mention the organization or person.

Dataset Owner

This is you by default. Only change it when you are adding data for someone else. Please keep the format the same as the account name, which is normally the same as the microsoft account name. E.g. "Ottow, Bouke".

Dataset Owner Email

Also filled in automatically. See Dataset Owner, only change when the owner is someone else. This email address is used for correspondence about this dataset.

Dataset Version

Mention here the version number or name when a dataset has versioning. It can also be a year or date for example for admin boundaries.

Security classification

The security classification is used for accessibility of the datasets through this catalog. Low means everyone can see the dataset in the catalog (also when not logged in). Normal means only logged in people can see the dataset and if possible a preview will be disclosed. High

The List of countries

Used in Project:

eg. Spain, India

Used in Product:

e.g Project Name

Data Quality

Expected Update Frequency:

e.g 10 days

Accuracy:

Accuracy of the Data

Consistency:

Consistency of the Data

Completeness:

Completeness of the Data

Date Added:

2022-06-27

The data license you select above only applies to the contents of any resource files that you add to this dataset. By submitting this form, you agree to release the metadata values that you enter into the form under the Open Database License.

* Required field

Next: Add Data

2. Create Dataset:

- d. When all the fields are filled in click on 'Next: Add Data'.

WHEN DO YOU USE THIS SCREEN

You use this when creating a new dataset.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

← → ↻ https://datacatalog.510.global/dataset/fda/resource/new

510 An initiative of the Netherlands Red Cross

Datasets About Search

Organisations / 510 / fda / Edit / Add New Resource

Add new resource

To add a new resource to an existing or newly created dataset you first have to select where the data(files) are stored. This is either through

- *External Link*, when the data is not in the 510 data platform but can be accessed online through a link (e.g. a google sheet);
- *Databases*, when the data is stored in one of the data bases in the 510 data platform;
- *Datalake*, when the data is stored in the Azure Data Lake Storage in the 510 data platform.

URL (External Link)

Put here the uri through which the data can be accessed.

Select database type (Databases)

Select here the type and server in the 510 data platform where the data is stored. This is either MySQL, Azure SQL or PostgreSQL.

Select database (Databases)

Select here the database (in the before selected server) where the data is stored. A popup will open that asks you to login to check if you actually have access to the database.

External Link > Databases > Datalake >

3a 3b 3c

3. Create Dataset:

- If the data is stored externally and accessible through a link click 'External Link'.
- If the data is stored on one of the 510 database servers click 'Databases'.
- If the data is stored in the 510 Azure Datalake Storage click 'Datalake'.

WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

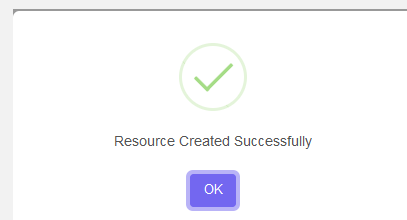
How to use the 510 Data Catalog

Adding a Dataset

The screenshot shows the 'Add New Resource' form in the 510 Data Catalog. The form is divided into two main sections: a left sidebar with instructions and a right main form area. The left sidebar contains three sections: 'Add new resource' (i), 'URL (External Link)' (iii), and 'Select database type (Databases)' (iii). The right main form area contains three tabs: 'External Link >', 'Databases >', and 'DataLake >'. The 'External Link >' tab is selected, showing a 'Resource with Link' form with fields for 'URL:', 'Name:', 'Description:', and 'Format:'. A 'Submit' button is at the bottom of this form. Callout 'ii' points to the 'Description' field, and callout 'iv' points to the 'Submit' button.

3a. External Link:

- i. Read general information about resources.
- ii. Fill in the information about the External Link.
- iii. Look up advice how to fill in the fields.
- iv. When everything is filled in click on 'Submit'.



WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

The screenshot shows the 'Add New Resource' page in the 510 Data Catalog. The page has a dark blue header with the 510 logo and 'An initiative of the Netherlands Red Cross'. The breadcrumb trail is 'Organisations / 510 / fda / Edit / Add New Resource'. There are three tabs: 'External Link >', 'Databases >' (selected), and 'Datalake >'. The 'Databases >' tab is highlighted with a yellow box and labeled 'ii'. The left sidebar contains three sections: 'Add new resource' (labeled 'i'), 'URL (External Link)' (labeled 'iii'), and 'Select database type (Databases)'. The 'Add new resource' section explains that data can be added via an external link, a database, or a datalake. The 'URL (External Link)' section asks for the URL. The 'Select database type (Databases)' section has a dropdown menu with 'Postgresql' selected. Below this is another dropdown menu labeled 'Select Database:' with 'Select Database' selected. There are also text input fields for 'Name:' and 'Description:'.

3b. Database:

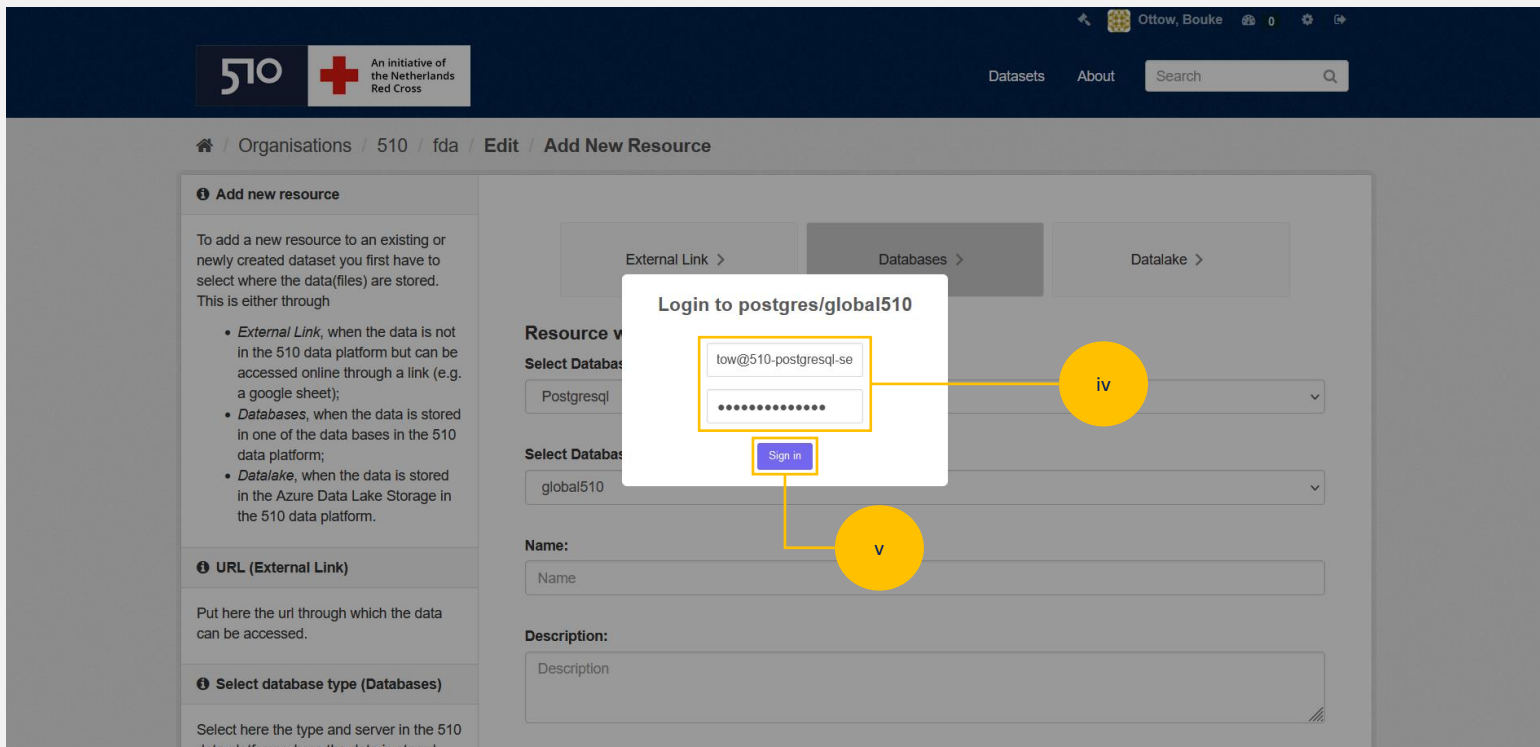
- i. Read general information about resources.
- ii. Select the type of the database where your dataset is stored and select the name of the database.
- iii. Look up advice how to fill in the fields.

WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

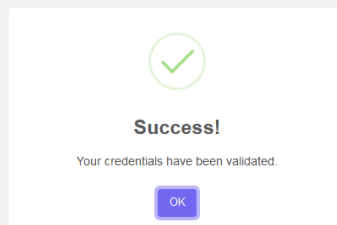
Adding a Dataset



3b. Database:

iv. Fill in your login to the database.

v. Click 'Sign in'. It will show the below pop-up if you have indeed access with that login.



WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

The screenshot shows the 'Add Dataset' form in the 510 Data Catalog. The form is divided into two main sections: a left sidebar with instructions and a right main form area. The left sidebar contains instructions for adding a dataset, including a note about the 'Data Lake' and a section for 'URL (External Link)'. The right main form area contains several fields: 'Select Database:' (dropdown menu), 'Select Schema:' (dropdown menu), 'Select Table:' (dropdown menu), 'No of Records:' (text input), 'No of Attributes:' (text input), 'Is Data Geo-Data?' (checkbox), 'Cell Size:' (text input), 'Temporal Resolution:' (text input), and 'Spatial Extent:' (text input). Annotations are present: 'vi' points to the 'Select Database:' dropdown, 'vii' points to the 'No of Records:' and 'No of Attributes:' fields, and 'viii' points to the 'Select Schema (Databases)' section in the left sidebar.

Annotations:

- vi: Select Database:
- vii: No of Records: and No of Attributes:
- viii: Select Schema (Databases)

3b. Database:

- vi. Select Schema and then the table of the dataset.
- vii. The number of records and attributes will be filled in automatically.
- viii. It will also recognize if the data is geo-data and will then fill in cell size (if it is raster) and spatial extent.

WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

Select here the schema (in the before selected database) where the data is stored.

Select Table (Databases)

Select here the table (in the before selected schema) which has the data.

Select Data from Datalake (Datalake)

Click the button and browse through the file structure in the data lake to select your data. Double click to go deeper in to a container or directory and click once to select the file or directory.

Geo-data (Databases & Datalake)

Check the *Is Data Geo-data?* checkbox to be able to (auto-) fill in the spatial metadata fields.

If the data is a table in a **postgres database** the box is automatically checked if the table has a geometry column and the fields are automatically populated. Depending if the data is raster data or vector data not all the fields might be applicable.

If the data is stored in the **data lake** you

Temporal Resolution

Spatial Extent:

(13565.4, 306846.198, 278026.09, 619232.555006162)

Temporal Extent:

Temporal Extent

Spatial Reference System:

28992

Name:

Name

Description:

Description

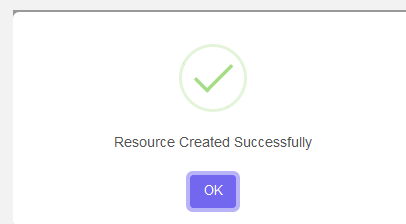
Format:

Database Table

Submit

3b. Database:

- ix. Fill in temporal resolution and extent and the name and description. Format is always Database Table.
- x. And finally click Submit.



WHEN DO YOU USE THIS SCREEN

You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

Add new resource

To add a new resource to an existing or newly created dataset you first have to select where the data(files) are stored. This is either through

- External Link**, when the data is not in the 510 data platform but can be accessed online through a link (e.g. a google sheet);
- Databases**, when the data is stored in one of the data bases in the 510 data platform;
- Datalake**, when the data is stored in the Azure Data Lake Storage in the 510 data platform.

URL (External Link)

Put here the url through which the data can be accessed.

Select database type (Databases)

Select here the type and server in the 510 data platform where the data is stored.

Add Resource from Datalake

Select Data from Datalake

Name:

Name

Description:

Description

3c. Datalake:

- Read general information about resources.
- Click select Data from Datalake and browse the datalake for your dataset. Double-click on a container or folder to open it, single select a file or folder to select it.
- Click on Select to select the file or folder.
- Look up advice how to fill in the fields.

*Double click to get data from Directory and Single Click to Select

Search for files	
..Back	
_\$azuretmpfolder\$	directory
dengue	directory
drought	directory
flood	directory
general	directory

Select

Records Per Page: 5

Total pages: 2

WHEN DO YOU USE THIS SCREEN

You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

The screenshot shows the 'Add Dataset' form in the 510 Data Catalog. The form is titled 'Select Data from Datalake'. It includes a sidebar with instructions and a main form area. Annotations are as follows:

- v**: Points to the 'Selected Directory/File' input field, which contains the path 'ibf/flood/floodscan/Uganda_1998-2019/1998'.
- vi**: Points to the 'Is Data Geo-Data?' checkbox, which is checked.
- vii**: Points to the 'Auto-fill Spatial Metadata' button.

The form also includes fields for 'Name' (1998), 'Base File (for Auto-fill)' (Choose file to generate spatial metadata from), 'Cell Size' (Spatial Resolution), 'Temporal Resolution' (Temporal Resolution), 'Spatial Extent' (Spatial Extent), and 'Temporal Extent' (Temporal Extent).

3c. Datalake:

- v.** Check the check-box next to 'Is Data Geo-Data?' only if it is geo-data.
- vi.** The application can then automatically extract the details, select here which file.
- vii.** Click the button to confirm. This is limited to datasets of 100 Mb and can take some time. Confirm by clicking Continue.

The modal dialog is titled 'Auto-fill spatial metadata?'. It contains the text: 'To auto-fill, we will download the file to our servers. This process may take time. Please reconsider if the file is big.' Below the text are two buttons: 'Continue' (highlighted in blue) and 'Cancel' (grey).

WHEN DO YOU USE THIS SCREEN
You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Adding a Dataset

Temporal Resolution

Spatial Extent:

[29.5, -1.5000000000000004, 35.1, 4.3]

Temporal Extent:

Spatial Reference System:

4326

Description:

This field is required

Description

Format:

image/tiff

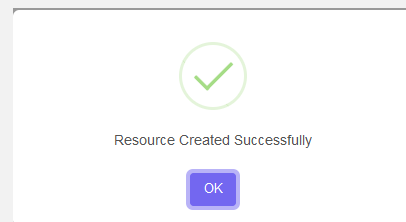
No of Files:

92

Submit

3c. Datalake:

- iix. Fill in the description. Format will be automatically extracted if possible, otherwise fill this in as well.
- ix. And finally click Submit.



WHEN DO YOU USE THIS SCREEN

You use this when selecting the dataset source.

510 Data Catalog

How to use the 510 Data Catalog

Edit a Dataset

The screenshot shows the 510 Data Catalog interface. The top navigation bar includes the 510 logo, the Red Cross logo, and the text 'An initiative of the Netherlands Red Cross'. The main header shows 'Organisations / 510 / test-dataset-for-manual'. The left sidebar contains a 'test-dataset-for-manual' section with a 'Followers' count of 0, a 'Follow' button, and an 'Organisation' section with a building icon and the text '510 There is no description for this organisation'. The main content area shows the dataset 'test-dataset-for-manual' with a 'HIGH' status. Below the dataset name are two data items: 'Veiligheidsregios' (The districts of the red cross) and 'aer_sfed_3s_20150105_v05r00.tif' (flood map for uganda). Each item has an 'Explore' button. At the bottom, there is an 'Additional Info' table.

Field	Value
Dataset name	test dataset for manual
Source	
Dataset Owner	Ottow, Bouke
Dataset Owner Email	BOttow@redcross.nl
Dataset Version	

4. Manage Dataset:

- Find the dataset you want to edit (see page 23 and further) and click 'Manage'

WHEN DO YOU USE THIS SCREEN
You use this when you want to edit a dataset

510 Data Catalog

How to use the 510 Data Catalog

Edit a Dataset

The screenshot shows the 'Edit' page for a dataset named 'test-dataset-for-manual' in the 510 Data Catalog. The page has a dark blue header with the 510 logo and navigation links. The main content area is divided into a sidebar and a main form. The sidebar on the left contains the dataset name, a 'Followers' count of 0, a 'Create Dataset' button, and a list of two entities: '1. the dataset' and '2. the resource'. The main form on the right has tabs for 'Edit metadata' and 'Resources'. Callout 'iii' points to the 'Resources' tab. The 'Edit metadata' tab is active, showing fields for 'Dataset name' (pre-filled with 'test dataset for manual'), 'URL' (pre-filled with 'datacatalog.510.global/dataset/test-dataset-for-manual'), 'Description' (with a placeholder 'eg. Some useful notes about the 510 data'), 'Tags' (with a placeholder 'eg. economy, mental health, government'), and 'License'. Callout 'ii' points to the 'Description' field. A yellow arrow points down from the 'Tags' field to the 'License' field.

4a. Manage Dataset:

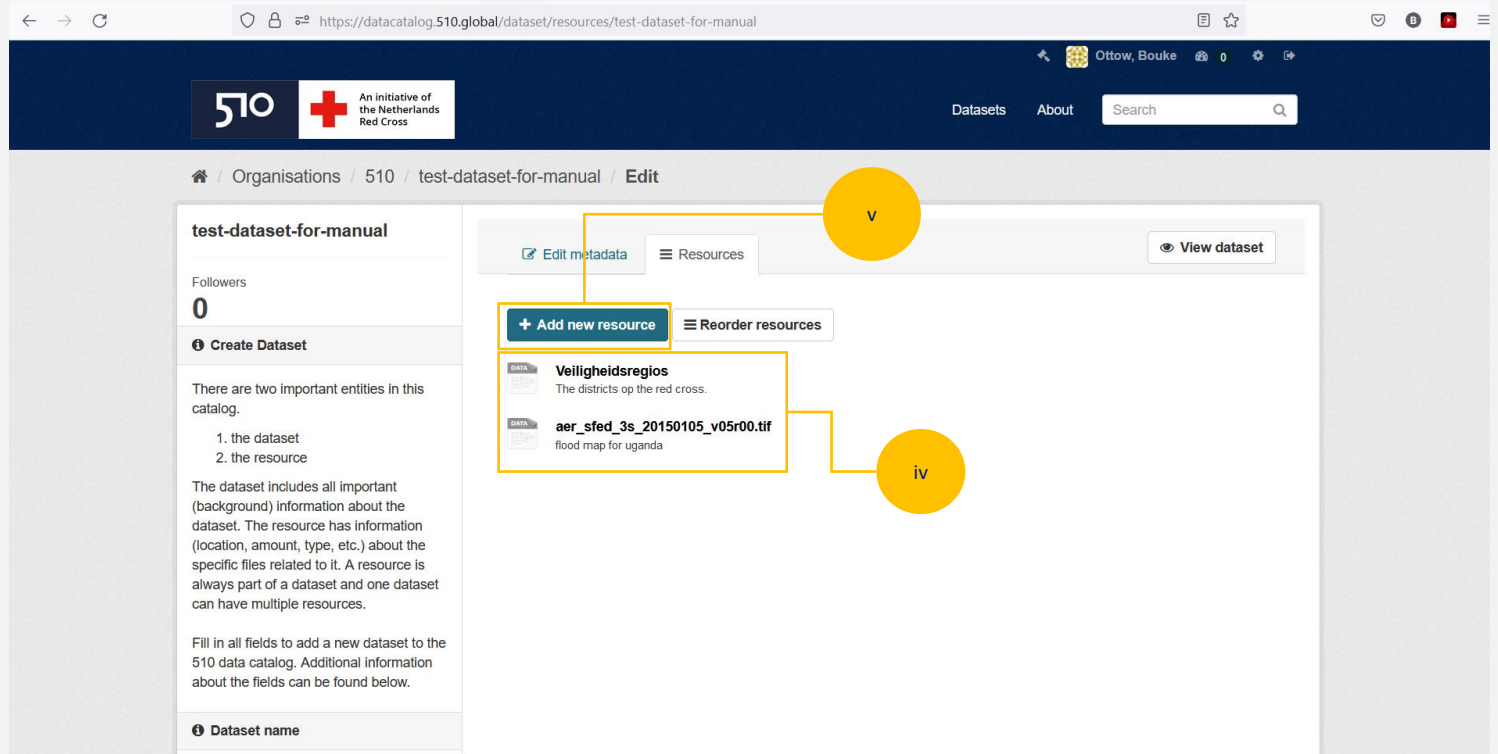
- ii. Edit here any of the metadata fields of the dataset.
- iii. Click 'Resources' to add or edit a resource.

WHEN DO YOU USE THIS SCREEN
You use this when you want to edit a dataset

510 Data Catalog

How to use the 510 Data Catalog

Edit a Dataset



4a. Manage Dataset:

- iv. Click on a resource to edit the metadata of the resource.
- v. Click 'Add new resource' to add another resource. See page 11 and further for help on selecting a resource.

WHEN DO YOU USE THIS SCREEN
You use this when you want to edit a dataset

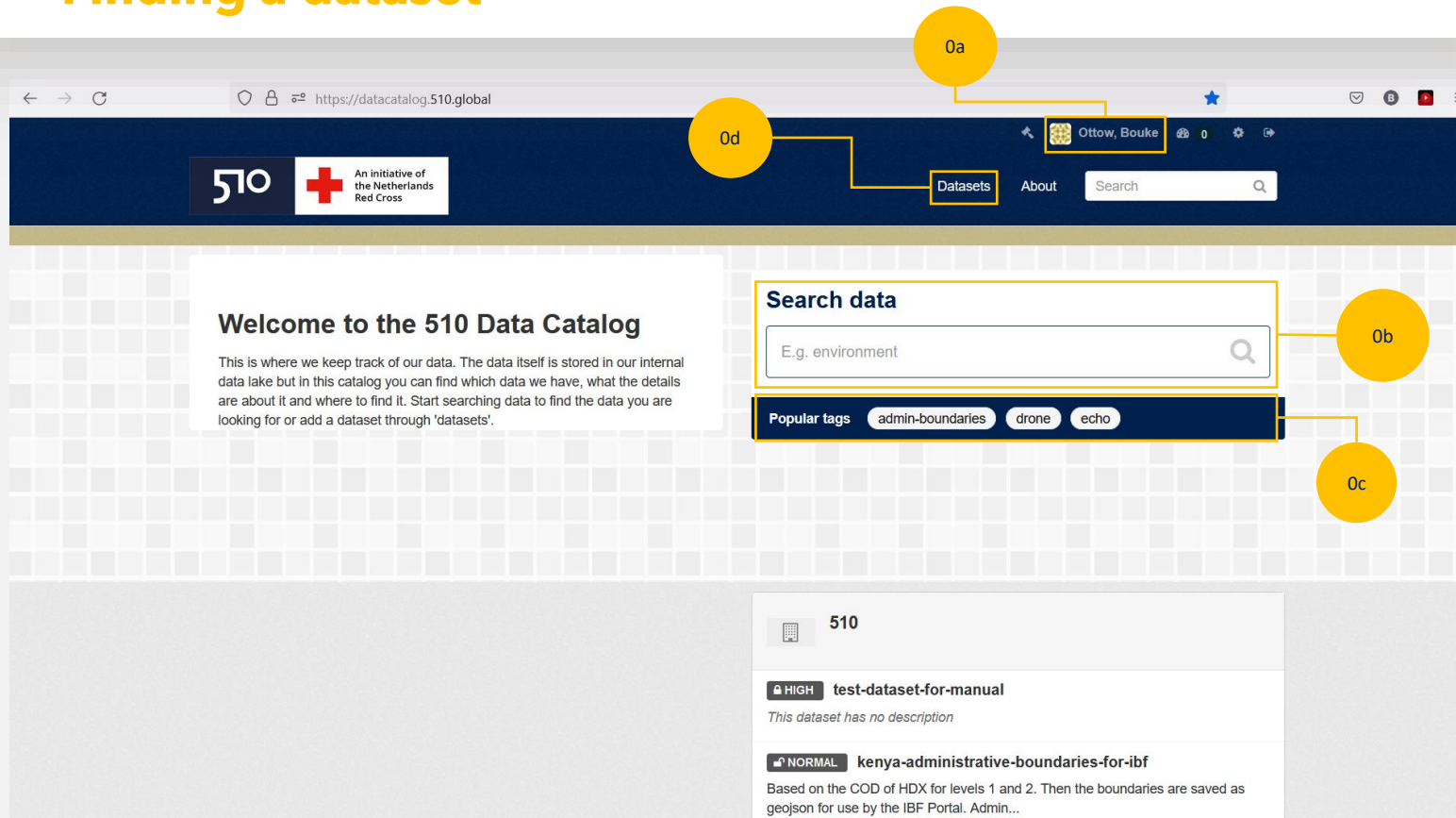
510 Data Catalog

Finding a dataset

510 Data Catalog

How to use the 510 Data Catalog

Finding a dataset



0. Home:

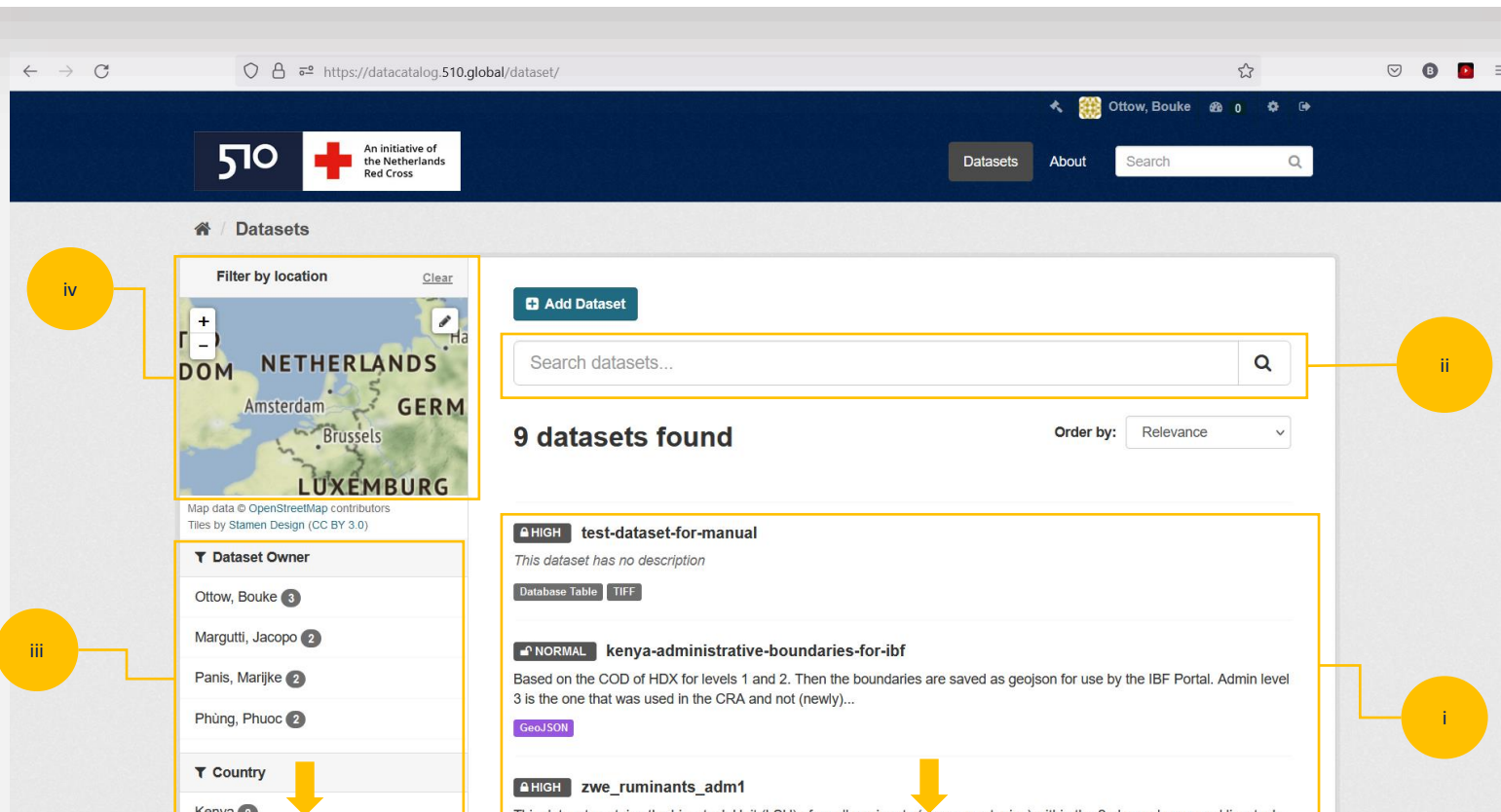
- a. Make sure you are logged in (see page 4 and 5).
- b. You can directly type in the search box to start searching for keywords.
- c. You can also directly click on a popular tag to filter datasets on a tag.
- d. Or you go to Datasets first for an overview of all the datasets (preferred route to start).

WHEN DO YOU USE THIS SCREEN
You use this when you search for a dataset

510 Data Catalog

How to use the 510 Data Catalog

Finding a dataset



0d. Datasets:

- i. You can select a dataset directly from the list.
- ii. You can search for keywords to make the list of datasets found smaller.
- iii. You can filter the dataset on Dataset Owner, Country, Dataset Quality, Security Classification, Project or Product.
- iv. You can also do a geo search by drawing a rectangle on a map returning all the datasets of which the first geo-data resource has an overlapping extent.

WHEN DO YOU USE THIS SCREEN
You use this when you search for a dataset

510 Data Catalog

How to use the 510 Data Catalog

Finding a dataset

The screenshot shows the 510 Data Catalog interface. The top navigation bar includes the 510 logo, a search bar, and links for 'Datasets' and 'About'. The breadcrumb trail indicates the current location: 'Organisations / 510 / test-dataset-for-manual'. The left sidebar displays the organization '510' with a 'Follow' button and social media links. The main content area has tabs for 'Dataset', 'Groups', and 'Activity Stream'. The 'Dataset' tab is selected, showing the details for 'test-dataset-for-manual'. The 'Data and Resources' section lists two datasets: 'Veiligheidsregio's' (The districts of the red cross) and 'aer_sfed_3s_20150105_v05r00.tif' (flood map for uganda). The 'Additional Info' section contains a table with the following data:

Field	Value
Dataset name	test dataset for manual
Source	
Dataset Owner	Ottow, Bouke
Dataset Owner Email	B.Ottow@redcross.nl
Dataset Version	

Annotations i.A and i.B point to the 'Dataset Owner Email' field and the 'aer_sfed_3s_20150105_v05r00.tif' dataset, respectively.

0d. i Dataset:

- A Here you find the information about the dataset.
- B Here you find the resources attached to the dataset, click on it to find out more.

WHEN DO YOU USE THIS SCREEN
You use this when you search for a dataset

510 Data Catalog

How to use the 510 Data Catalog

Finding a dataset

The screenshot shows the 510 Data Catalog interface. The URL is <https://datacatalog.510.global/dataset/test-dataset-for-manual/resource/a909dfc-5beb-4c87-89ad-13203222b75>. The page title is 'Veiligheidsregios'. The main content area displays the DB Connection String for a postgres Database: `postgresql://<username>:<password>@510-postgresql-server.postgres.database.azure.com/global510`. Below this, it says 'The districts op the red cross.' and a note: 'Note: A preview cannot be shown for data that is in the Datalake or a database.' On the left, there is a 'Resources' section with a link to 'Veiligheidsregios' and a resource ID 'aer_sfed_3s_20150105_v...'. On the right, there is an 'Additional Information' table.

Additional Information	
Date Last Edited	2022-03-04
Created	2022-03-04
Format	Database Table
License	License not specified
Database Connection Type	postgres
Database Host	510-postgresql-server.postgres.database.azure.com
Database Name	global510
Schema Name	nederland
Table Name	veiligheidsregios

Annotations on the screenshot:

- i.E: Points to the 'Request Data' button.
- i.C: Points to the DB Connection String.
- i.F: Points to the 'Resources' section.
- i.D: Points to the 'Additional Information' table.

0d. i Dataset:

- C This is the connection string to connect to the dataset through code.
- D Here you find the information about the resource and the location where it is stored.
- E If the resource is stored in a location where you don't have access to you can request access by sending an email to the Dataset Owner. This button starts the email.
- F Here you can browse to the other resources linked to this dataset.

WHEN DO YOU USE THIS SCREEN
You use this when you search for a dataset

510 Data Catalog

Frequently Asked Questions

510 Data Catalog

How to use the 510 Data Catalog

Frequently Asked Questions

Questions people ask a lot: