Hand-in-Hand Data Catalogue Workflow Documentation

2023

Hand-in-Hand Data Catalogue Workflow Documentation

This document highlights the workflow for the datasets management and publication on the FAO Data Catalogue portal.

Table of Contents

THE HAND-IN-HAND INITIATIVE

Hand-in-Hand is an evidence-based, country-led and country-owned initiative of the Food and Agriculture Organization of the United Nations (FAO) to accelerate agricultural transformation and sustainable rural development to eradicate poverty (SDG 1) and end hunger and all forms of malnutrition (SDG 2). You can read all about it here.

One of its initiatives is to deploy sophisticated tools, including advanced geospatial modelling and analytics, to identify the biggest opportunities to raise the incomes and reduce the inequities and vulnerabilities of the rural poor, who constitute the vast majority of the world's poor. For this initiative to be realized there is the need for a geospatial platform.

The Hand-in-Hand Geospatial Platform

The Hand-in-Hand Geospatial Platform is used to expose data available in FAO and to make them accessible for public use. Over 20 FAO units, across multiple domains, have contributed to the platform with data on Soil, Land, Water, Climate, Fisheries, Livestock, Crops, Forestry, Trade, Social and Economics, etc. Data have also been sourced by FAO partners and public data providers across the UN and NGOs, academia, private sector and space agencies. So far, we have assembled a million geospatial layers, thousands of statistics series, including food security and environmental indicators, to assist all stakeholders working in the food and agriculture sector planning their target interventions. We organize and publish the datasets and related metadata records using our Geospatial Data Catalogue portal.

DATA CATALOGUE PORTAL

The FAO Data Catalogue portal manages collections of data and metadata that are exposed through the Hand-in-Hand Geospatial Platform as well as any other FAO systems using the same technology. It is built based on the Comprehensive Knowledge Archive Network (CKAN), which is a powerful data management system for making open data websites. CKAN allows national and local governments, research institutions, private companies and other organizations to share their data with the public. Users who browse CKAN catalogues can find the data they need and preview it using maps, graphs and tables. CKAN is a free, open source and customizable software.

The FAO Data Catalogue portal can be accessed here. It also exposes its content to other information systems through public APIs.

The Data Catalogue portal (Figure 1) consists of three main components – **Datasets**, **Organizations** and **Groups**.
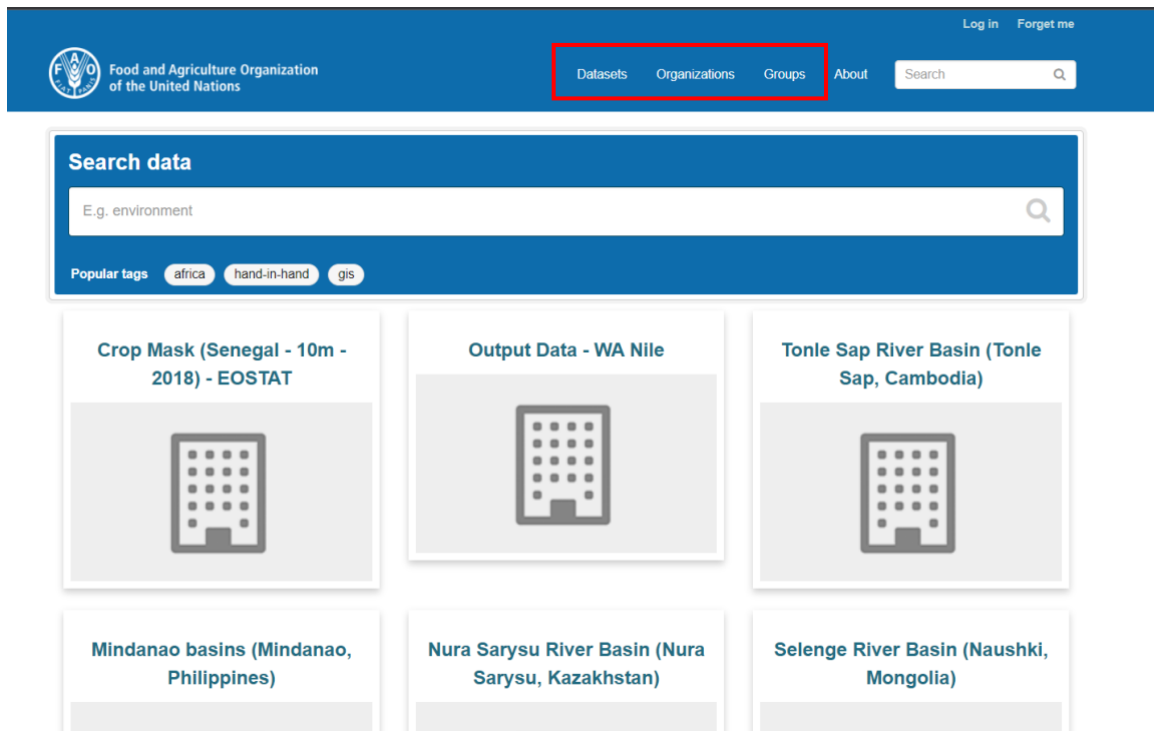
Figure 1. FAO Data Catalogue portal homepage

Datasets

A Dataset is a parcel of data that includes the resource with the related metadata information. A data resource can be a CSV or Excel spreadsheet, XML file, PDF document, image file, and geospatial files.

A dataset can contain multiple resources, representing the same data in different formats or for different years. CKAN can store the resources internally, or store them as links to external sources on the web.

Individual datasets have their own fixed URL and they are what users get from the catalog when searching for data.

Datasets by default are private and can be accessible by the public only when they are published. In this way, data providers have time to finalize their datasets, making sure the information provided is complete and correct before publishing it. Datasets are easier to find when they are well described with accurate metadata details (Figure 2).

Figure 2. Datasets search results page

Organisations

An organization is used to create, manage, and publish collections of datasets and to manage users (Figure 3). Users can have different roles within an organization, depending on their level of authorisation to create, edit and publish information.

Figure 3. Organizations search results page

Groups

We can use groups to organize and manage collections of already existing datasets. This could be to catalogue datasets for a particular project or team, or on a particular theme or a geographical area, or as a very simple way to help users find and published datasets. Currently, we use groups to organize datasets under countries (Figure 4).

Figure 4. Groups search results page

User Registration and Login

All users without registration have access to any published datasets available in the Data Catalog but, registered users assigned to one or more organizations can have access to private information and may get special privileges to create new datasets, manage content and groups within their own organisation(s).

This section provides help for new users on how to register and login to the Data Catalogue portal. It covers registration, login as well as clearing login credentials when logged in on an _untrusted_ browser.

Note that, registering to the platform does not automatically give you permissions or privileges, which are assigned by an administrator of organisation or group. Indeed, you have to inform your HiH focal point after your registration in order to get the right permissions. Learn more about permissions in here

User Registration

A user who will use the FAO Data Catalogue to publish datasets and related metadata needs to register with the platform. The registration process requests to provide the organization with the identification details. Please follow the two step-registration below to learn how to become a registered user of the FAO catalog. The "**Log in**" button on the top-right corner of the homepage (Figure 5) brings you to the first registration page.
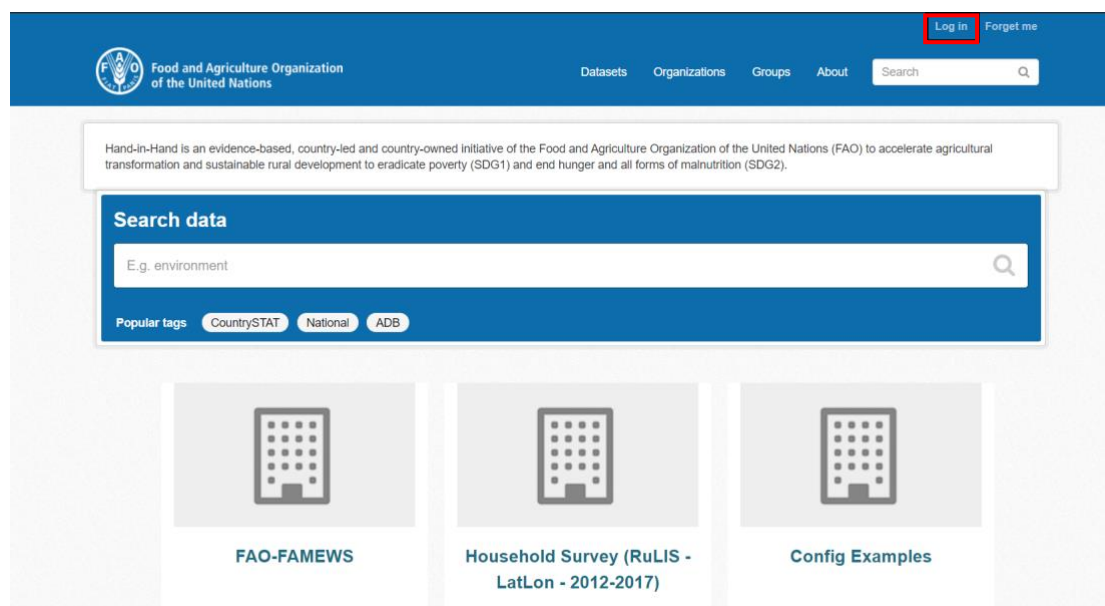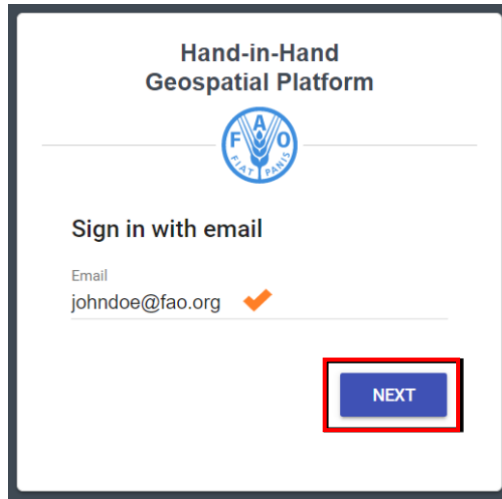
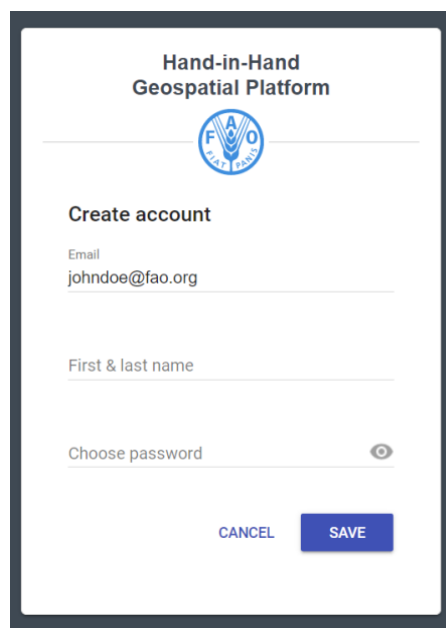

Figure 5. Log in button on the homepage

1: Sign in with your email address (Figure 6). Please make sure to enter your email in **lowercase** and then click on the "**NEXT**" button.

Figure 6. Registration page

2: Fill in your full name and password. You can make the password visible to be sure it is correct, by clicking on the eye icon, and then click on "**SAVE**".

Figure 7. Registration page with identification details

Once the registration is complete, the platform will create your account and automatically log you in. You will be redirected to the homepage of the Data Catalogue portal with your details (Figure 8). After creating your account, you may inform the administrator of the organisation you wish to belong to.
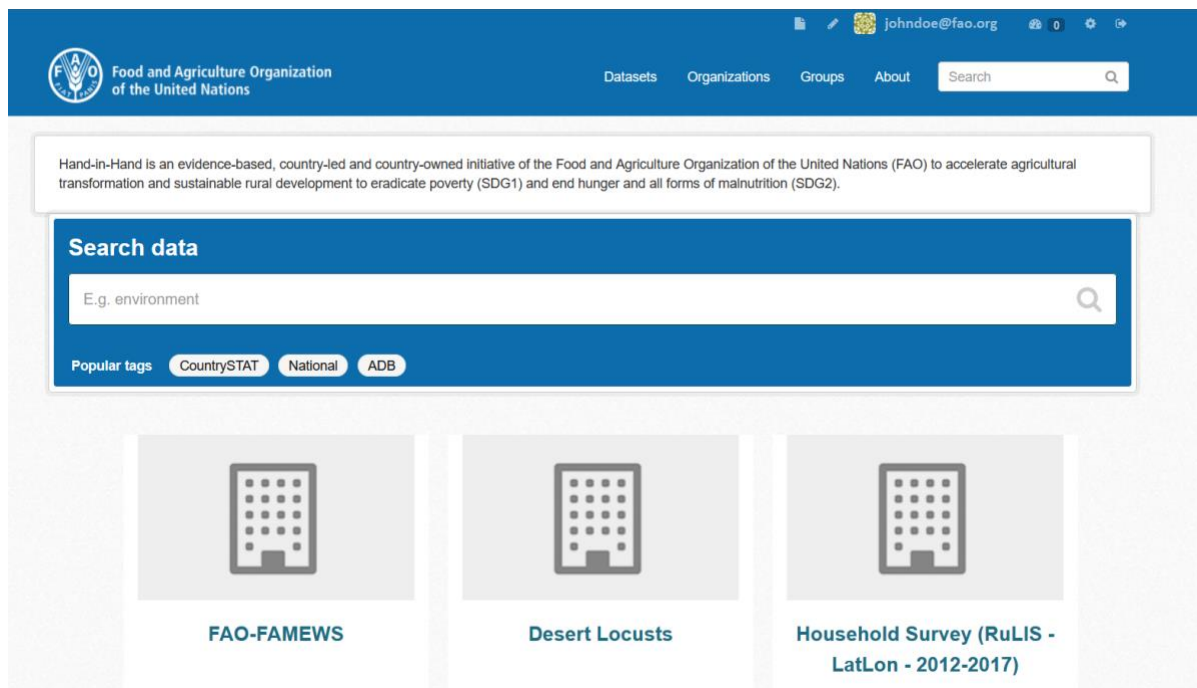

Figure 8. Homepage of an authenticated user

User Login

A user already registered can login in to the Data Catalogue portal anytime using the "**Log in**" function (Figure 5). The system may automatically identify your account and log you in if you are on the same browser used for the registration, and then it may bring you to the page on which you were before logging in.

In case you are using an Incognito page rather than a different browser or any computer that is not your personal one, the login button will bring you to the first registration page (Figure 6) requesting your email again. Please remind to enter your address in lowercase.

Once you click on the **"NEXT"** button, the system is able to identify your email and ask for your password. You may finally complete your authentication by clicking on "**SIGN IN**".

Figure 9. Authentication page

Forgotten Password/ Password Recovery

Should you forget your password, there is no cause for alarm. Please follow the steps below to learn how easily you may recover your account.

1: Go to the authentication page (Figure 9) by following the instructions illustrated in the previous paragraph and then click on the **Trouble signing in?** link. That should bring you to the page from where you may send the request for resetting your password (Figure 10). Please verify that your email address is correct and then click on "**SEND**".

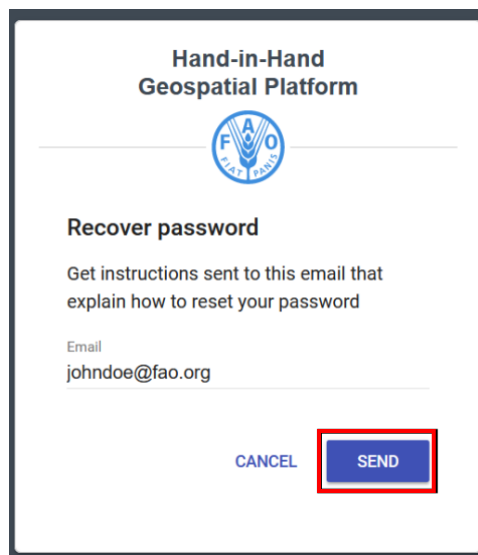Figure 10. Recover password.

You will receive an email with the subject **Reset your password for data.apps.fao.org.** The email will provide you with a recovery link.

2: Use this link to open the page (Figure 11) where you may enter your new password and then click on the "**SAVE**" button.
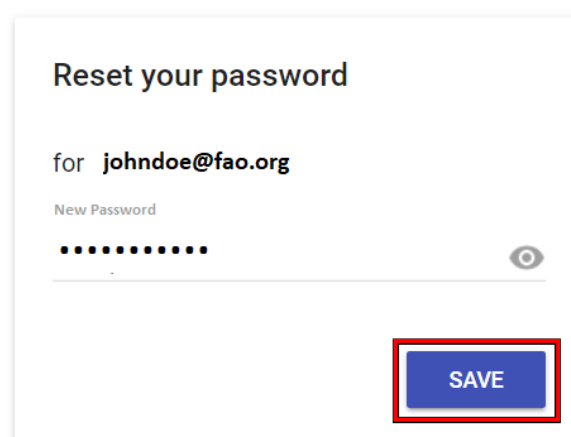

Figure 11. Reset password

3: Click on "**Log in**" from the homepage and sign in with the new password.

Forget Me

With this section you can learn how to clear or delete your account credentials from a browser. You may use this function when you no longer want to use a specific browser to log in to the portal. It could be when

- you logged in with a computer that belongs to someone else or is shared
- you have changed your personal computer.

> Note: We advise that you log in using a Private Window or Incognito mode on any computer that is not your personal one and close the browser when you are done.

Step 1: Log out by clicking on the last icon at the at the top right corner of the homepage (Figure 12)
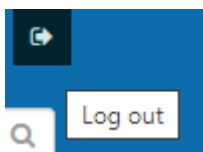

Figure 12. Log out icon

Step 2: Use the "**Forget me**" link (Figure 13) you may see appearing next to the "**Log in**" link to clear or delete your account credentials from a browser.

 Figure 13. Forget me link

DATASETS PUBLISHING WORKFLOW

The creation of a dataset and its following publication on the Data Catalog is organized in three main sections: 1- Adding a new dataset, 2- Reviewing content, and 3- Publishing a dataset. The full process requires different roles and privileges assigned to the users, which are illustrated here below.

Permissions

Four different user profiles are available in the CKAN platform:

- **Visitor**: unregistered user of the platform, who can:

    - browse the platform

    - view all **public** or **published** datasets

    - **register** to be a member of the platform.

- **Member** : registered user of the platform, who can:

    - browse the platform

    - view all **public** or **published** datasets

    - view **private** or **unpublished** datasets in assigned organization

- **Editor**: registered user of the platform with special editing privileges, who can

    - view all **public** or **published** datasets

    - view **private** or **unpublished** datasets in assigned organizations

    - **create**, **modify**, **delete** or **move** the datasets in owned organizations

- **Admin**: registered user of the platform with administration privileges within an organization or group, who can

    - view all **public** or **published** datasets

    - view **private** or **unpublished** datasets in assigned organizations

    - **create, modify**, **delete** or **move** the datasets between owned organizations

    - **manage** the datasets visibility by **publishing** or **unpublishing** them.

    - **manage** owned organizations by adding registered users (as *Editor* or *Admin*)

Adding a dataset

This section defines the workflow of the dataset creation, from the new metadata generation with related resources associated to the data preview configuration. It will requires some steps to be completed, all being essential to proceed further.

As we mentioned in the previous paragraph, in order to add and edit datasets in the Data Catalog, you need to be a member of at least one of its organizations, with the assigned role of Editor or Administrator.

There are two ways to add a new dataset in the catalogue: from the **Datasets** link and from the **Organizations** link, both at the top of the home page.

Option 1: From the Datasets link, you will be immediately redirected to the "**+ Add Dataset**" button to start a new metadata form for documenting your data (Figure 14).
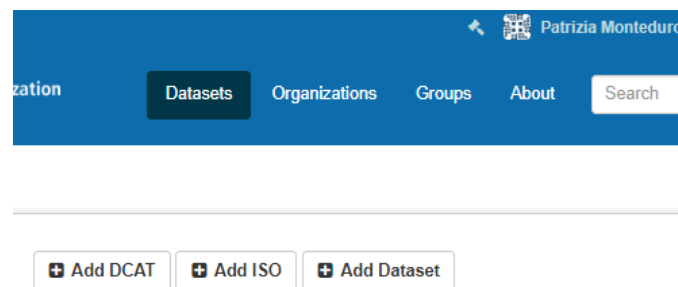

Figure 14. Links to the metadata standards available from the Datasets page.

Option 2: Alternatively, navigating the **Organization** page, you may select the organization that should own your new dataset and then, if you are a member of this organization, you can select the "**+ Add Dataset**" button above the search box (Figure 15).
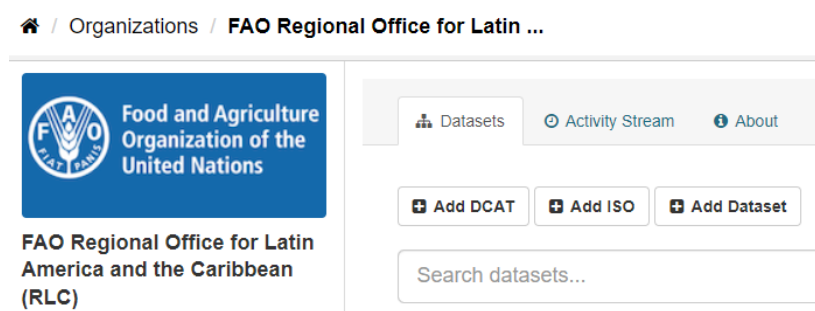

Figure 15. Links to the metadata standards types available from the **Organization** page

The following steps may guide you on the overall process.

1. Fill in the metadata form (Figure 16) with information about your data using the following metadata elements:
   - **Title**: This title will be unique across CKAN, and used by the system to generate the URL. Thus, make sure it is brief but specific.
   - **Description**: You can add a longer description of the dataset here, including information such as where the data is from and any information that people will need to know when using the data.
   - **Tags**: here you may add tags that will help people find the data and link it with other related data. Examples could be "population", "food security", "Africa". Hit the <return> key between tags. If you enter a tag wrongly, you can use its delete button to remove it before saving the dataset.
   - **License**: It is important to include license information so that people know how they can use the data. This field should be a drop-down box. If you need to use a license not on the list, contact your site administrator.
   - **Organization**: If you are a member of multiple organizations, this drop-down will enable you to choose which one should own the dataset.
   - **Source / Version**:
   - **Author/ Author e-mail**: The name and e-mail address of the person or organization responsible for producing the data.
   - **Maintainer / maintainer e-mail**: If necessary, details for a second person responsible for the data.
   - **Custom fields**: If you want the dataset to have another field, you can add the field name and value here. E.g. "Year of publication".

Note: By default, all the above metadata fields are optional except the title. However, it is good practice to include, at the minimum, a short description, the Author details as point of contact and, if possible, the license information. You should ensure that you choose the correct organization for the dataset, particularly if you have opened the metadata form from the Dataset link.

Figure 16. Metadata form

2. When you have filled in the metadata form, select the "**Next: Add Data**" button (Figure 17) in order to complete the information with details on the associated resource. You may link any type of resource to the form: documents, tables, graphics, images, Web services, etc...

Figure 17. Metadata form – Add Data button

3. Depending on the data location, you may upload your file directly from your computer or add a link to an external resource. Fill out the basic information to describe the resource and use the autocomplete feature to assign a format to it (Figure 18).

Figure 18. Data resource description form

4. Provide here the information about the resource. It is good practice to add:
   - Name – a name for this resource, e.g. "Population density 2011, CSV". Different resources in the dataset should have different names.
   - Description – a short description of the resource.
   - Format – the file format of the resource, e.g. CSV (comma-separated values), XLS, JSON, PDF, etc.

5. Click on the "**Save and add another**" button to save the current resource and add another one, otherwise click on Finish to complete the process (Figure 18).

The metadata record with the associated resources is successfully created but it will be accessible only by the members of that organisation until its publication.

Creating views for datasets

When resources are added to datasets, they can be visualized on the catalogue portal. The preview feature supports (views) mainly geospatial data, displaying the resource content in a map viewer.. Below is an example of how to create a map view for a geospatial data.

To create a view:

1. Locate the dataset and navigate to the resource (Figure 19)
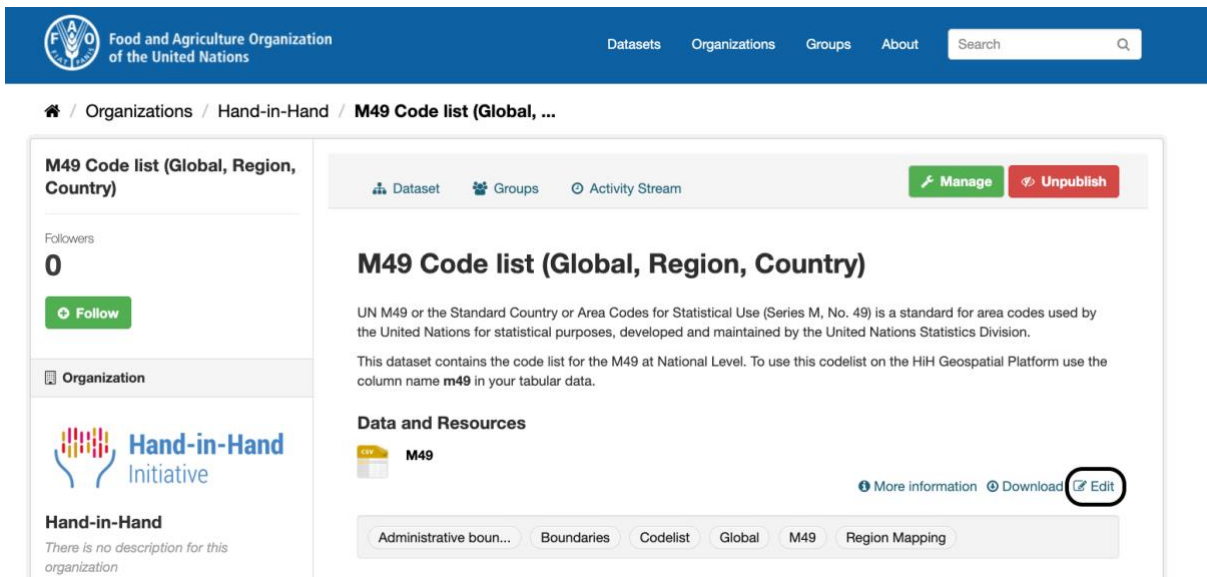2. Click on the "**Edit**" link

Figure 19. Sample dataset

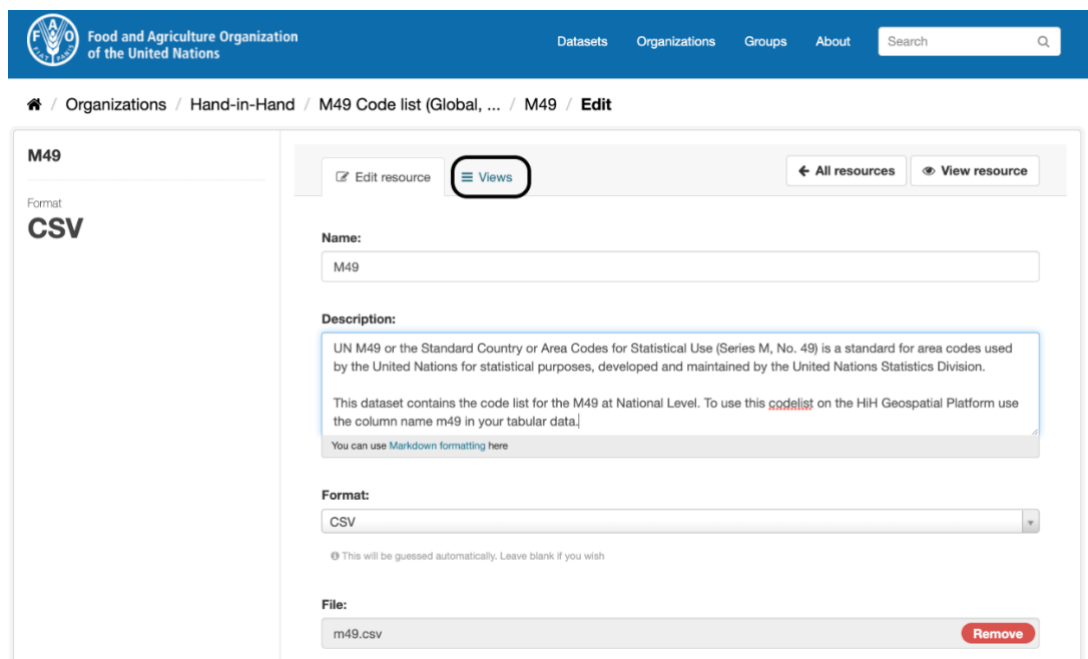3. Switch to the Views tab (Figure 20).

Figure 20. Sample dataset in editing mode

4. Click on the + New view drop down button and select Map (see Figure 21)
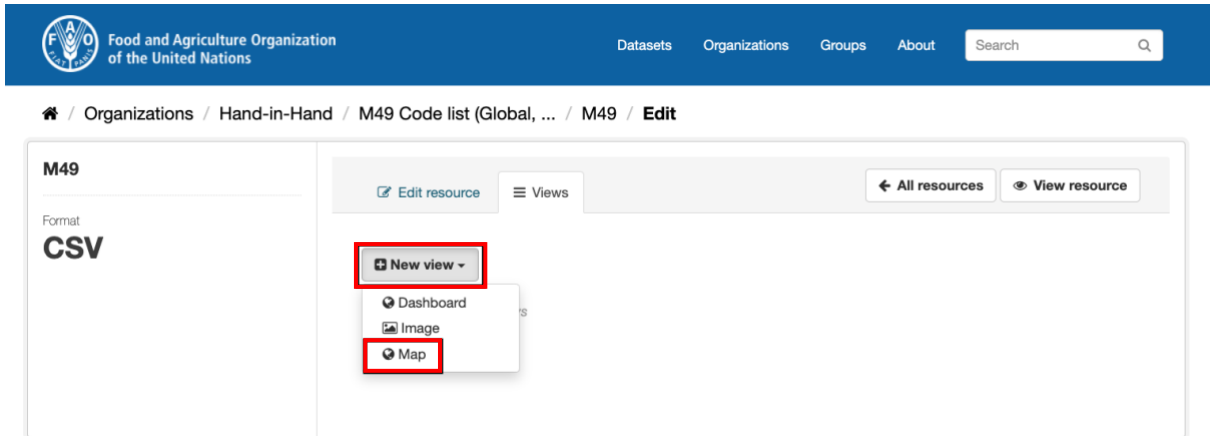
Figure 21. Views page.

5. You may navigate through the preloaded resource configuration (Figure 22) and add more root properties, if needed then click on the "**Add**" button to save the changes.

Figure 22. Resource configuration page in editing mode

6. Finally, click on the "**View resource**" button (Figure 23) to display your data on top of a global map provider by the map viewer (Figure 24) integrated within the Data Catalog.
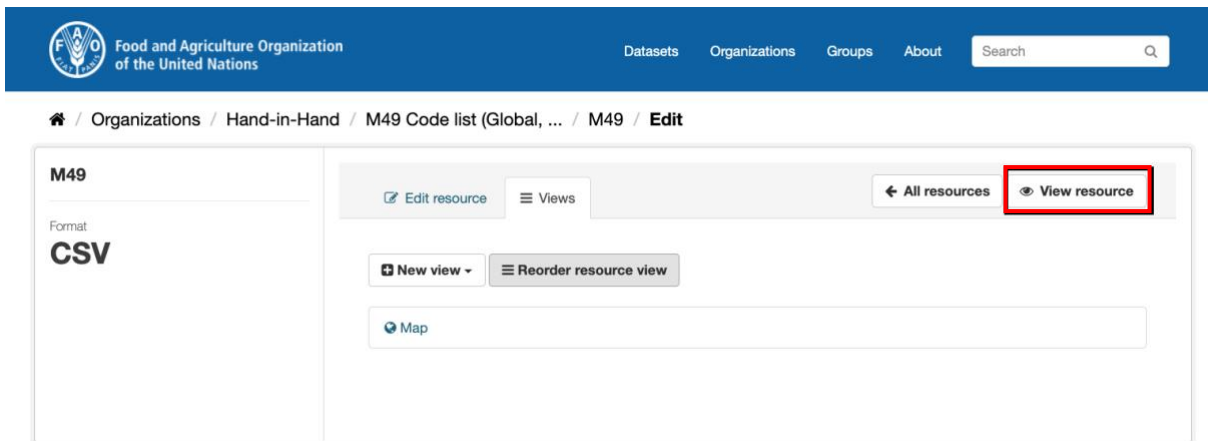
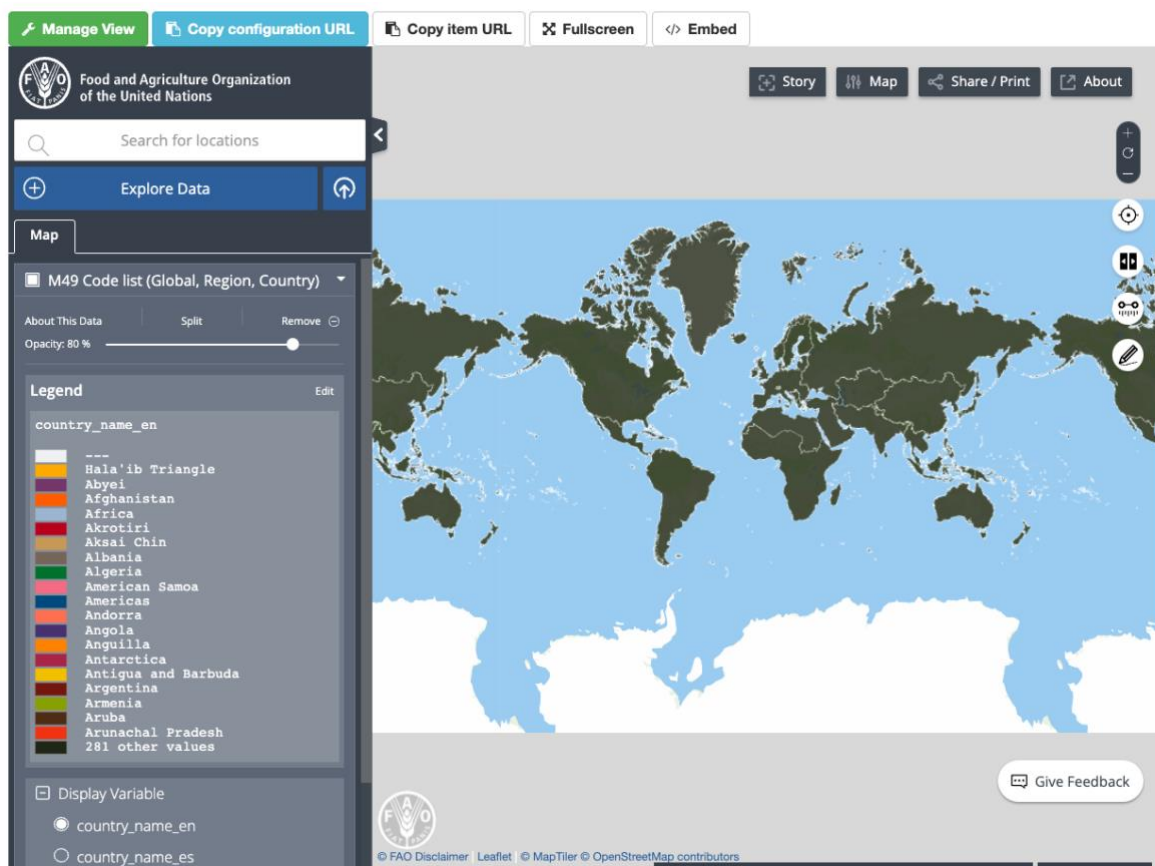Figure 23. View page with View resource highlighted

Figure 24. Sample dataset displayed in the map viewer

Reviewing a dataset

This is the last step of the dataset publication workflow, where the metadata and data views are reviewed by the editor and examined by the organization admin for the final check before making the new information available to the public. Here below are few points that are important to include in the checklist.

Ensure that:

- the names of the dataset and resources are without errors (See naming conventions)
- a description is added to the dataset and to resources.
- the right License is selected for the dataset and resources
- an author and/or maintainer is added
- the appropriate tags have been added to the dataset
- the dataset belongs to the right organization, in the case where you have access to more than one organization

Publish Datasets

By default, when a new dataset record is created, it is private and can be accessed only by registered members of the organization it belongs to. To make the dataset accessible to the public, it needs to be published and, as previously mentioned; only the administrator of the organisation can do it. To complete the dataset publication, the Admin needs to follow the steps below.

1. Search the catalog to locate the unpublished dataset. It comes in the search result page with a "**Private**" icon next to the dataset title (Figure 25).

Figure 25. Private dataset within the search result page.

2. Click on the dataset title to display the full content and review the metadata details with the attached resources (See Reviewing a dataset).
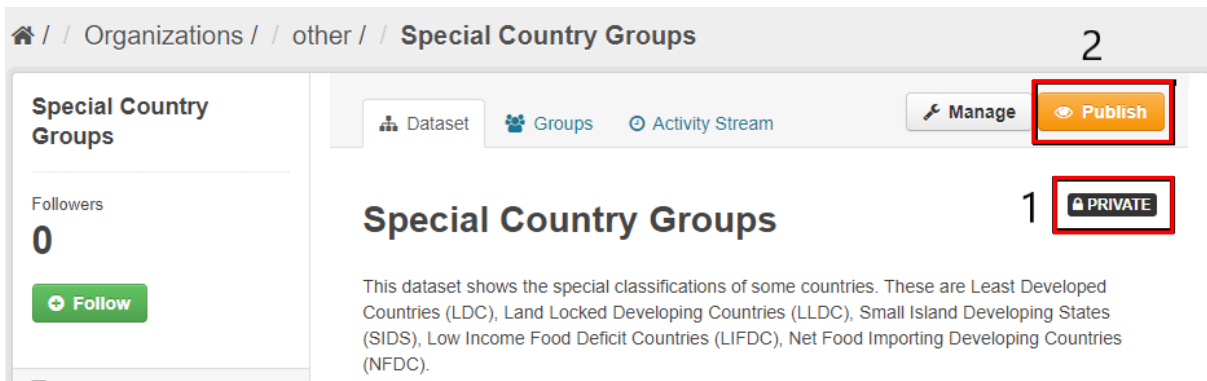3. Click on the "**Publish**" button to make the dataset visible to everyone (Figure 26).

Figure 26. Private dataset in view mode

Please note that the dataset can be put back to private any time by clicking on the Unpublish button highlighted in the image below (Figure 27)
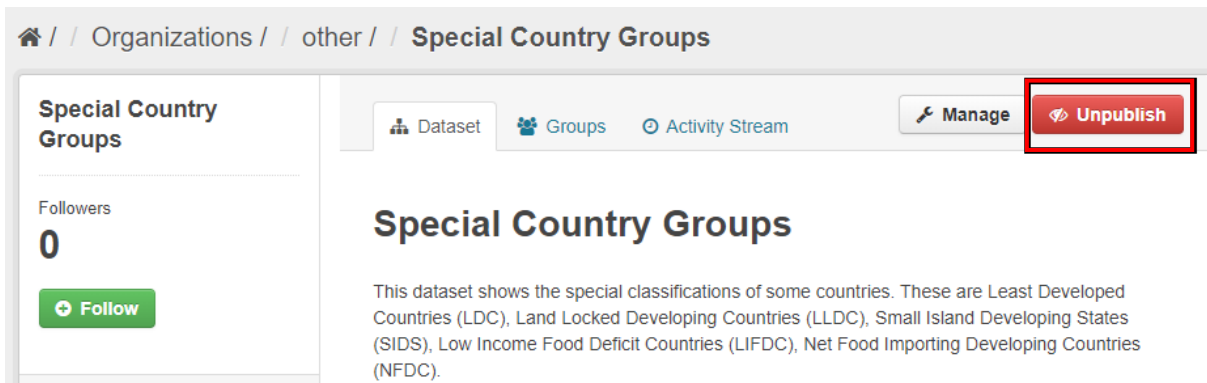

Figure 27. Public dataset in view mode

Workflow Diagram

The diagram below (Figure 28) highlights the main steps of the whole workflow of datasets publication, from the dataset creation to the resource view configuration, and from the dataset revision to the final publication. Please note that the core of the full process is the Admin of the data organization, who has the responsibility to make sure that the information published may guarantee comprehensibility and accuracy as well as accountability, data ownership and due acknowledgments.
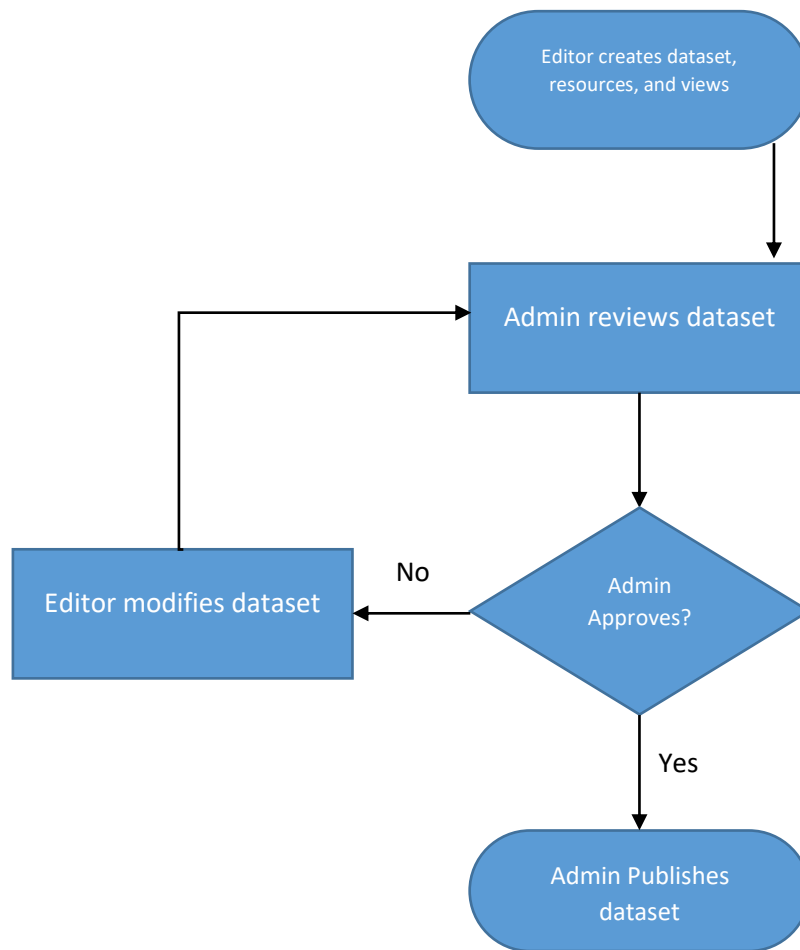
```
        ┌─────────────────────────┐
        │  Editor creates dataset,│
        │   resources, and views  │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │   Admin reviews dataset │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐        ◇ Admin
        │ Editor modifies dataset │◄── No── Approves?
        └─────────────────────────┘        
                                            │ Yes
                                            ▼
                                   ┌─────────────────┐
                                   │ Admin Publishes │
                                   │     dataset     │
                                   └─────────────────┘
```

Figure 28. Diagram of the datasets publication workflow

Dataset Naming Conventions – Best practices

Naming datasets on the catalogue portal is very important. The dataset name (title) must be consistent with the name and/or subject of the resource so that it will make it easier for users to retrieve it when searching for data.

> Note that the system gives a unique name for every datasets and thus, it is impossible to have two datasets with the same name. When creating a new dataset and the name already exists, the system will notify and prevent you from proceeding until you change the name.

Here below are a few suggestions on how to name your datasets.

- Keep the name as short and concise as possible (Max 10 words).
    - *Crop Production*
    - *Fertilizer Import Quantity*

    Any acronym, when available, must be placed in brackets or in dash signs.

    - *Statistics (AQUASTAT)*
    - *Global Agro-Ecological Zones - GAEZ 2015*

- Add dimensions (location, time frame, unit of measure) if applicable in brackets.
    - *Crop Harvest (Global National)*
    - *Major Markets (Ecuador – Lat Lon - 2022)*

    When using brackets, use a *comma* to separate elements of the same dimension and *dash* to distinguish between dimensions, according to the order: (*Area, Country – Time dimension – Resolution*)

    - *Evaporation (Zankalon, Egypt - Annual - 30m)*

- Avoid
    - *using only abbreviations as names*
    - *very long names for datasets and resources*

    A well-formatted title should also avoid special characters that may be misinterpreted by computer systems, such as *slash, back-slash, percentage symbol, quotation marks*, etc. In the following example, *slash* (in bullet point 1) could be easily replaced by *dash* (in bullet point 2):

    - *Global Agro-Ecological Zones - GAEZ v4 (Global/10km);*
    - *Global Agro-Ecological Zones - GAEZ v4 (Global - 10km)*

ORGANISATION MANAGEMENT

Every organisation needs at least a user with Admin permissions to manage it. Only users with Admin role in an organisation can add other registered users to it. This section will cover the organisation management and users registration to an organisation.
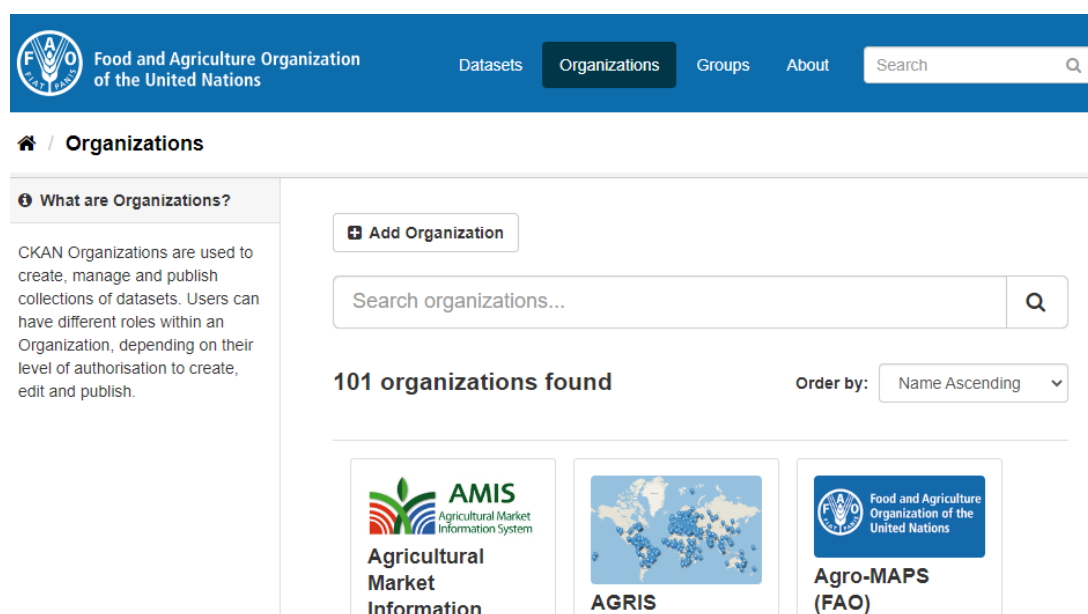
Managing your organisation

Organisations by default can only be created by the Data Catalog administrator with super admin permissions. You need to make a request to fao-data@fao.org to have a new organisation created. Once the request is made and the organization is created, the owner(s) is(are) assigned to the organisation as Admin(s). With this role, the users can update the details of their organisations as well as manage other registered users.

Edit Organisation

To edit your organisation details, please follow the steps below:

1. Login to the Data Catalog portal (Figure 5)
2. Navigate to your organisation from the "**Organization**" page (Figure 29)

Figure 29. Search organization page

3. Click on the "**Manage**" button (Figure 30) to display the organization record in editing mode
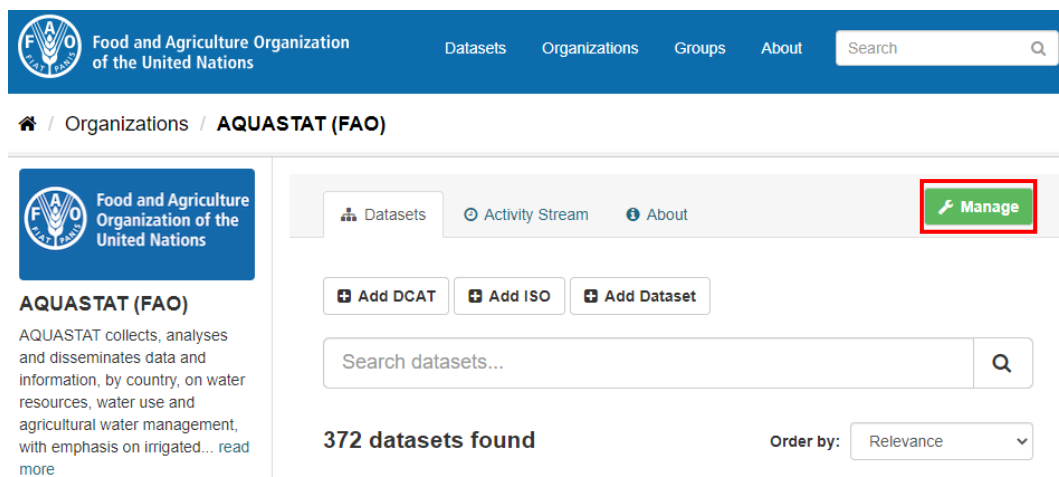

Figure 30. Sample of an Organization in view mode: "Manage" button highlighted

4. Edit the record (Figure 31). A short description of the organization is suggested, as well as the contact details. You can add an image to the organization - by either uploading a local file or linking an external resource.
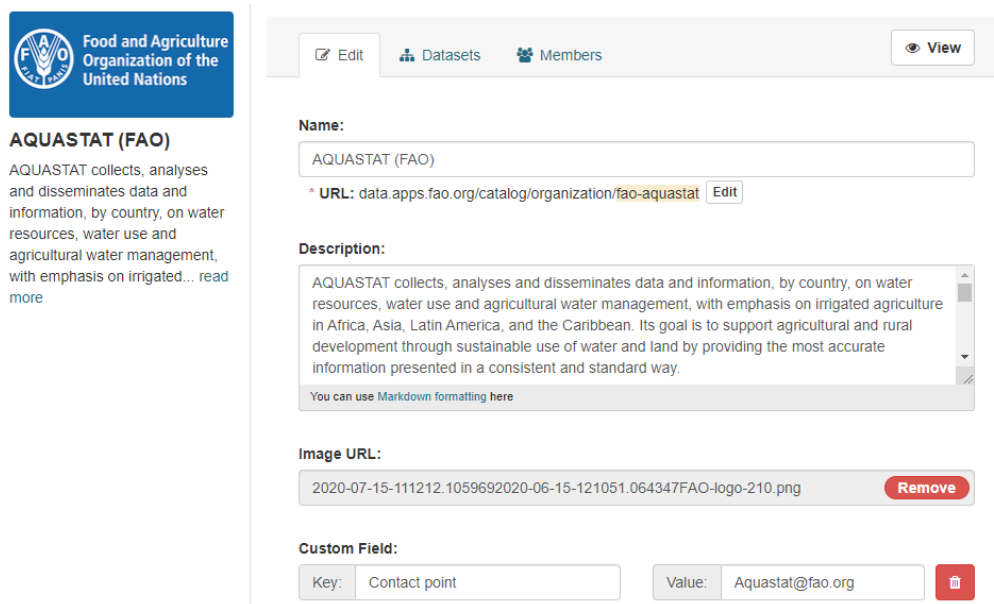

Figure 31. Sample of an Organization in editing mode

The image will be then shown in the metadata view on the Hand-in-Hand platform, for every datasets that belong to that organisation (Figure 32).
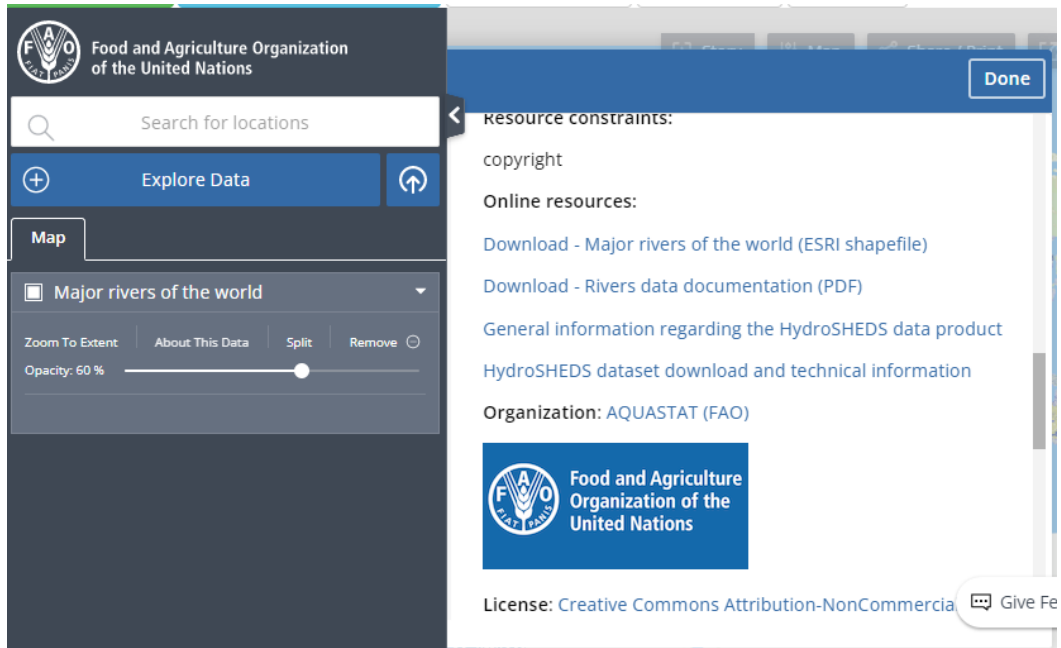
Figure 32. Sample of a metadata view in the Hand-in-Hand platform

5. Click on the "**Update Organisation**" to save the changes in the Organization page (Figure 33).

Figure 33. Organization in editing mode: Update Organization button highlighted

User Management in Organizations

This section defines how users in organizations are managed on the platform. Organizations act as data owners, with members who add, edit, review, and publish datasets. As already mentioned earlier in this document, in every organization there are three kinds of users – Admin, Editor and Member (Figure 34). Regarding users management within an organisation, only the Admin of that organization can do so. The Admin can,

- Add registered users to his/her assigned organization
- Assign, change or revoke user permissions within the organization.

Summary of User Profiles and Roles

Role	Datasets	Organization	Group
Admin	Add, edit, delete, publish and unpublish datasets	Add, edit, delete datasets to assigned organization	Add or remove dataset to assigned group
		Add members to assigned organization	Add members to assigned group
		Manage roles of members in assigned organization	
Editor	Add, edit, delete datasets	Add, edit, delete datasets to assigned organization	N/A
Member	View all published datasets	View all private datasets of assigned organizations.	View all published datasets
Visitor	View published datasets		View published datasets

Figure 34. User profiles and roles

Add User to Organization

In order to add a new user to an organization, its administrator can be guided by the following steps:

1. Navigate to the assigned organization (Figure 29).
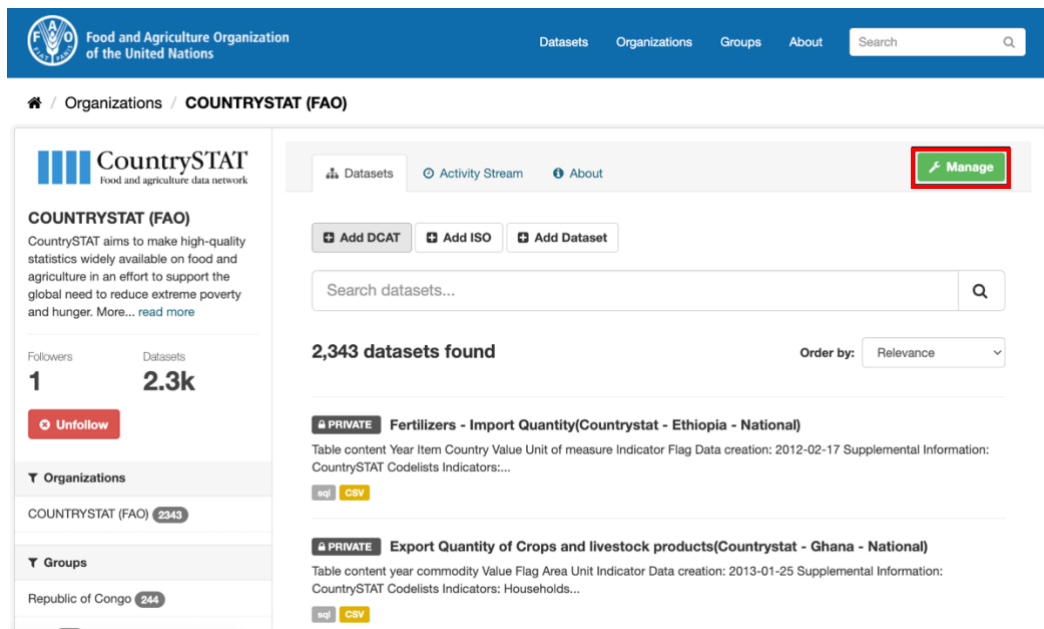2. Click on the "**Manage**" button (Figure 35)

Figure 35. Sample of Organization in view mode: "Manage" button highlighted

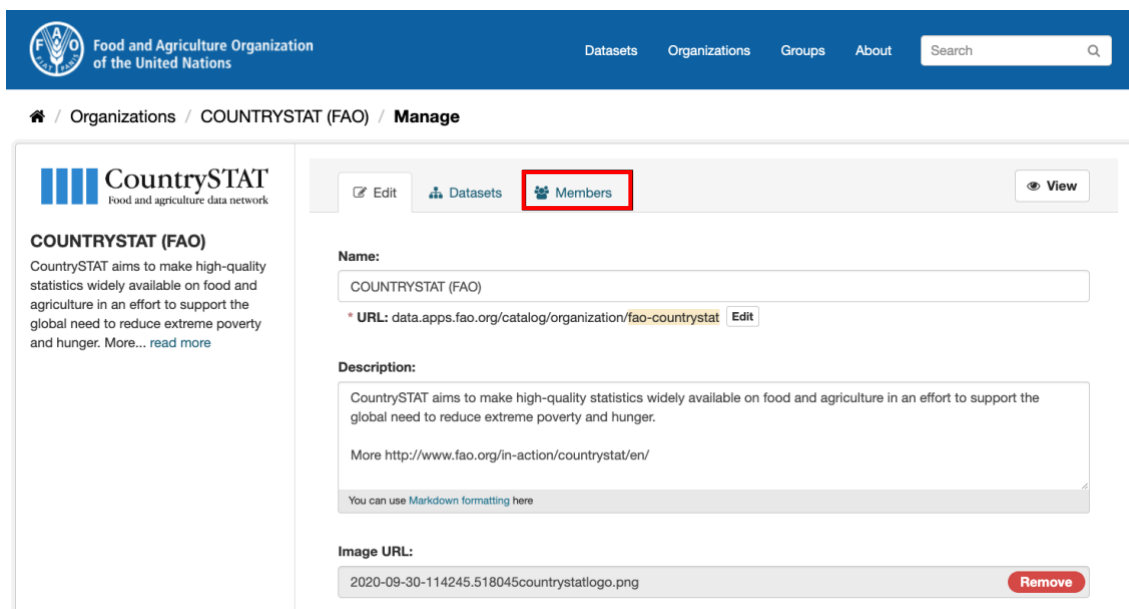3. Click on "**Members**" section (Figure 36)

Figure 36. Organization in editing mode: "Members" button highlighted

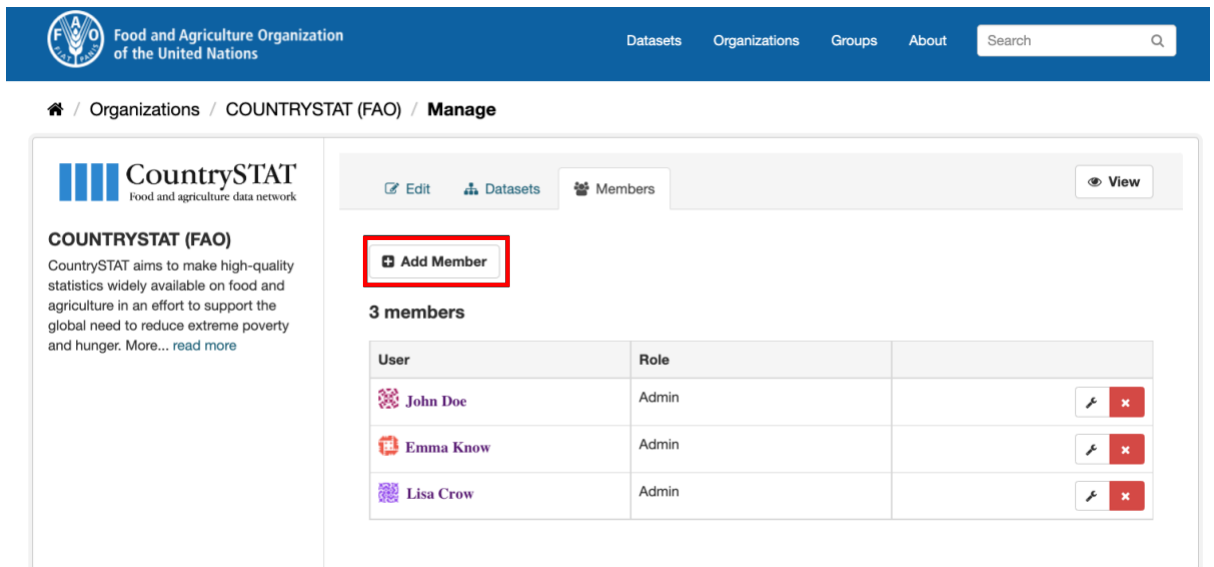4. Click on "**Add Member**" button (Figure 37)

Figure 37. "Members" page: "Add Member" button highlighted

5. Use the autocomplete feature (1) to add a registered user by username. Assign appropriate permission (2) and finally click on "**Add Member**" button (3) to complete the registration (Figure 38).
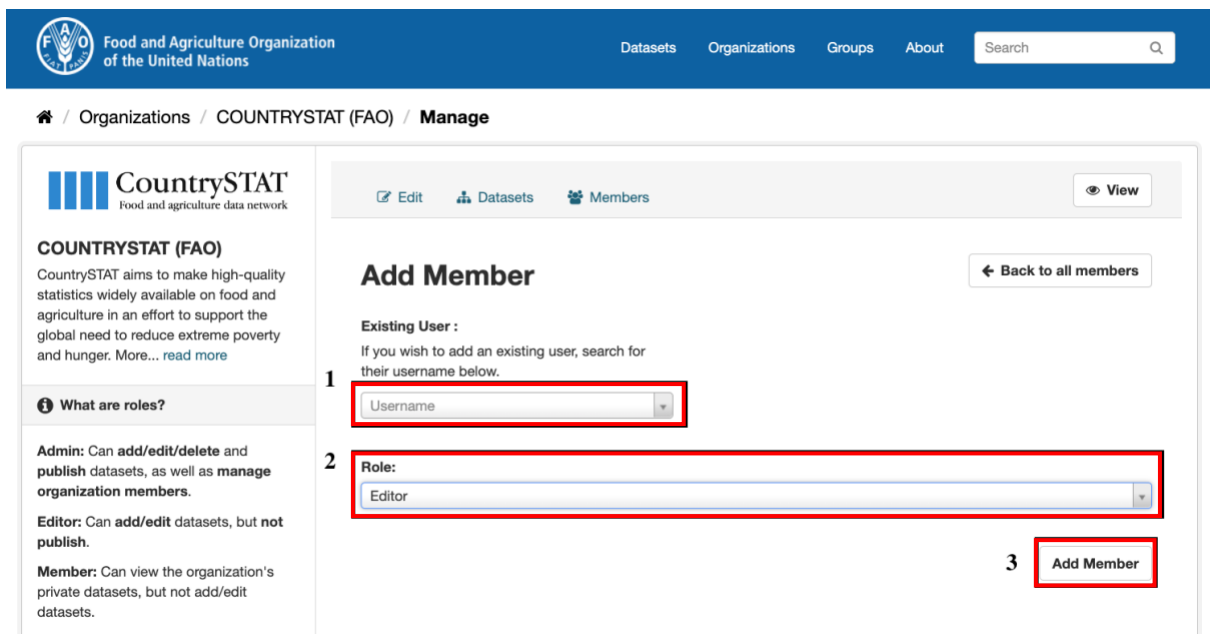

Figure 38. "Add Member" page

All maps in this publication have been created using shapefiles from the United Nations.
Source: FAO Hand-in-Hand Geospatial Platform. 2023. Map geodata [shapefiles]. New York, USA, United Nations.
The boundaries and names shown and the designations used on these map(s) do not imply the expression of any opinion concerning the delimitation of its frontiers and boundaries. Dashed lines on maps represent approximate border lines for which there may not yet be full agreement.