**Data Catalog User Guide**

***Ministry of Health, eHealth Africa***

***December 2017***

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Author** | **Date** | **Note** |
| 0.1 | Qing Gong | 2017.12.29 | draft |
| 0.2 | Qing Gong | 2018.02.27 | Updated based on discussions with Mike Mulbah |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## 

## 

## 

Table of Content

[Organizations](#_7ub39m919vyc)

[Create Organization](#_cgozsx2f1s4e)

[Edit Organization](#_961szaavywth)

[View Organization](#_qnbxgfrnduke)

[Datasets](#_6senoj3wazlk)

[Categories of Datasets](#_srtshtjcjvj0)

[Create Dataset](#_2bc5eu19cq1p)

[Enter Basic Information for the Dataset](#_t7jiynx06pse)

[Add Resources for Dataset](#_r54dxee9fba)

[Edit Dataset](#_7hzorbfa1msd)

[View Dataset](#_e5sna7li5ixb)

[Groups](#_kezah88s4j5a)

[Create Group](#_a6h87z3x3t0j)

[Edit Group](#_1aiqs1pn4zuz)

[Assign Datasets to Groups](#_gm5l6ihbpo2w)

[View Group](#_fes6n2gbin6c)

[User Roles and Access Control Rules](#_viiei7kivxf)

[Extensions and References](#_xt2201a5ux29)

## 

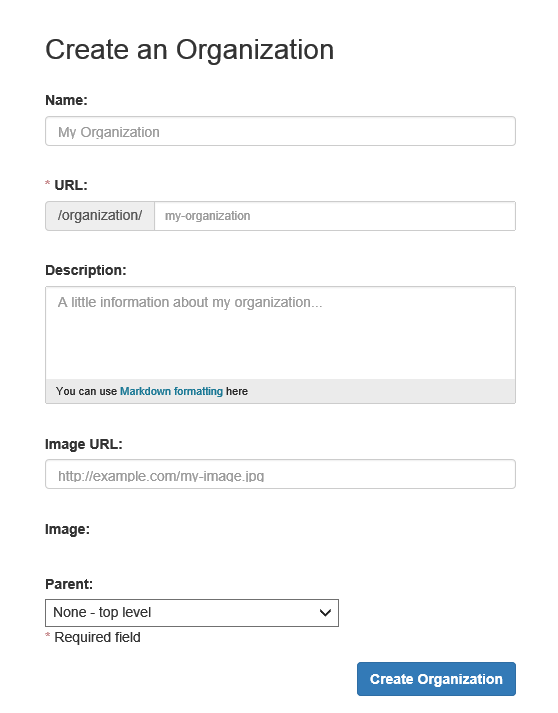
## 

## Organizations

### Create Organization

Since a dataset must belong to an organization, the belonging organization must be added before a dataset can be added.

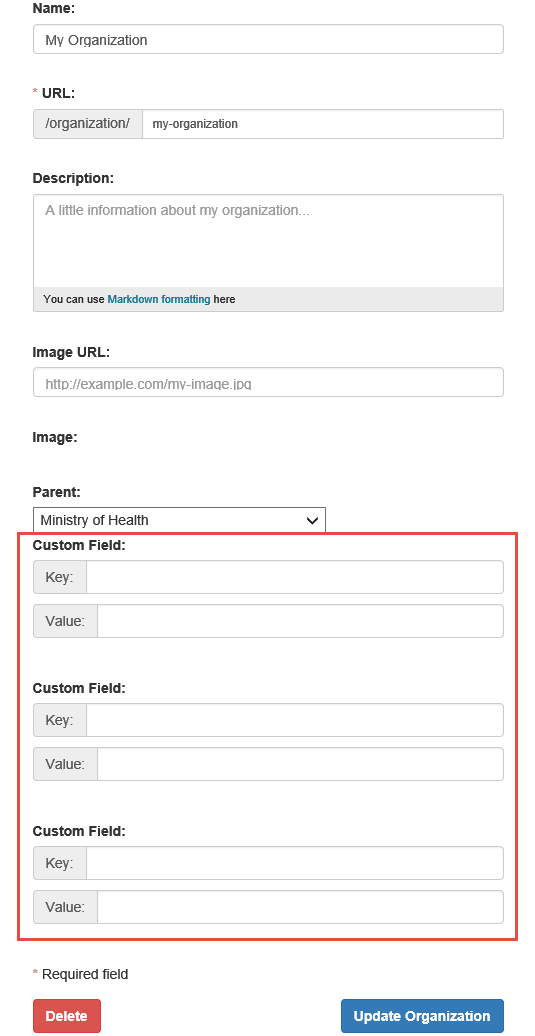
To create an organization, user shall navigate to *Organizations -> Add Organization*.



* **Name**: the name of the organizational unit in MoH, for example, Research Unit, Department of Policy, Planning & Research. The names of the organizations shall be same as the names of organizational units defined in the Organizational Structure of MoH.
* **URL**: the value entered in this field defines the URL address of this organizational unit. The name of the organization in lower case shall be entered here with “-” connecting each letter. For example, if the name of this organization is “Research Unit”, the value entered here shall be “research-unit”.
* **Description**: a brief description of the organization shall be entered here, including xxx?
* **Image URL**: an image/logo of the organization can be provided. If the image is available online, user can enter the URL address that leads to the image.
* **Image**: by clicking on the blank area below “Image”, a window will be opened where user can upload the image from his/her local computer.
* **Parent**: the parent organization of the current organization can be selected here. If the current organization is Human Resource Unit and it belongs to the Division of Administration. “Division of Administration” shall be selected here. All organizations that are created in the data catalog will be selectable from the dropdown list.

After user clicks on “Create Organization”, the organization will be created, and user will be directed to the page of the organization.

Apart from the already entered information, there are some other information that shall be provided for an organization. To do so, user shall click on the “Manage” button at the top right corner of the page. Three Custom Fields become available.



Depending on the situation, three of the following four fields shall be added. For example, for the organization Ministry of Health, Organization Website, Contact Phone Number, and Contact Email shall be provided. For the departments and divisions, Head of Organization, Contact Phone Number, and Contact Email shall be provided.

* **Organization Website**: the official website of the organization.
* **Head of Organization**: the name of the person who is responsible for the organization,
* **Contact Phone Number**: phone number of the person who is responsible for the organization.
* **Contact Email**: email address of the person who is responsible for the organization.

The name of the field shall be entered as “Key” while the content to be entered for the filled shall be entered as “Value”. For example, if the Head of Organization is Jon Snow, user shall enter “Head of Organization” as “Key” and “Jon Snow” as “Value” in a custom field.

After the custom fields are added, user can click on the “Update Organization” button to save the changes.

### Edit Organization

To edit the basic information of an organization, user can navigate to:

* Organizations *-> select an organization -> Manage -> Edit tab*

Three custom fields will appear at the first time when user tries to edit the organization. Unlike editing dataset, no more custom field will appear when user tries to edit the organization again.

To add members to an organization, user can navigate to:

* Organizations *-> select an organization -> Manage -> Members tab -> Add Member*

### View Organization

To view an organization, user can navigate to *Organizations -> select an organization*.

* “Dataset” tab: the “Datasets” tab will be displayed by default, where user can see all the datasets that belong to the organization. If user check the “Include Sub-Organizations” checkbox, datasets belong to all the sub organizations of the current organization will also be displayed.
* “Activity Stream” tab: all updates made to the organization will be listed under this tab.
* “About” tab: all the additional information about the organization added through the Custom Fields will be displayed here.

## Datasets

### Categories of Datasets

Three types of datasets are identified for the Data Catalog of MoH:

* Dataset: refers to data that are stored in the database of a HIS subsystem.
* Excel Tools: refers to data that are stored in Excel tools.
* Resources: refers to other types of information that are electronic or paper based documents, images, or videos.

### Create Dataset

There are two ways a dataset can be added.

* *Datasets -> Add Dataset*
* *Organizations -> select an organization -> Add Dataset*

#### Enter Basic Information for the Dataset

After user clicks on the “Add Dataset” button, the Create Dataset page will be displayed. On this page, a list of fields will be available by default. The user shall fill in some of them, leave some of them empty, and add some additional fields in accordance with the category of the dataset to be added.

The following fields already exist by default and shall be filled for dataset of all categories.

* **Title**: a short descriptive phrase or sentence that describes what the dataset is about. If the dataset comes from a system, the name of the system shall be provided, with the abbreviation of the system in the bracket, e.g. Community Based Information System (CBIS). When two datasets have the same name, the name of its belonging organization can be indicated in the title, e.g. Ebola Virus Disease Database - NRL. If the dataset is a resource, the name of the resource shall be provide.
* **URL**: the value entered in this field defines the URL address of this dataset. The title of the dataset in lower case shall be entered here with “-” connecting each letter. For example, if the title of this dataset is “My Dataset”, the value entered here shall be “my-dataset”.
* **Description**: a more detailed description of the dataset shall be entered here.
* **Tags**: users are free to enter related keywords as tags for the dataset. (any rules?)
* **License**: the license defines how the dataset viewer can use the information included in this dataset. For example, the license of a public dataset could be, “Open Data Commons Attribution License” shall be selected. (Need to define this) Before a decision is made, “License not specified” shall be chosen in this field.
* **Organization**: the organization this dataset belongs to shall be selected. When more than one organizational unit is involved in a dataset, the one that owns and maintains the dataset shall be selected. When a dataset is owned by one organization and maintained by another, the organization who owns the dataset shall be selected. For example, CBIS is owned by Community Health Service unit, and managed by M & E unit, in this case, the Community Health Service unit shall be chosen as its organization.
* **Visibility**: two options are available, public and private. When “Public” is selected, anyone who has access to the data catalog will be able to see the dataset. When “Private” is selected, only users who are members of the organization this dataset belongs to can see the dataset.
* **Version**: keeps track of the versions of the dataset in the data catalog. Note that this is not a sophisticated version control system, but to give an idea of the version of the dataset. When the dataset is added for the first time, the version shall be version 0.1. When modifications are made to the information of this dataset in the data catalog, the version number shall be increased as 0.2, 0.3, etc. When the dataset comes to a final version, the version can be changed to version 1.0.
* **Author**: the name of the user who is creating this dataset shall be entered here.
* **Author Email**: the email address of the author shall be entered here. (Question: who shall be the author? The person or the organization?)

The following fields shall be added manually and be filled for dataset of all categories. To add a field manually, the “Custom Field” will be used. The name of the field shall be entered as “Key” while the content to be entered for the filled shall be entered as “Value”. For example, if the contact of a dataset is John, the user shall enter “Contact” as “Key” and “John” as “Value” in a custom field. Note that by default, only three custom fields are available. After these three custom fields are filled, the user can first save the dataset, and come back again to edit the dataset. Once the edit page of this dataset is opened, a new custom field will be added.

* **Abbreviation**:if the dataset has an abbreviation that is widely used and referred to within MoH, it shall be added here. For example, CBIS, LMIS, iHRIS.
* **Contact**: the name of the contact person of this dataset. The contact person shall be someone who is responsible for and familiar with the dataset, for example, the HIS subsystem lead, the organizational unit lead, the data manager, etc. This person will be contacted when data viewers have any questions about this dataset. (Question: contact from the owner organization or the maintaining organization?)
* **Contact Email**: the email address of the contact person shall be entered here.
* **Contact Phone No.**: the phone number of the contact person shall be entered here.
* **Contact Position**: the position of the contact person shall be entered here.

The following fields shall be added manually and be filled for dataset that belongs to the dataset category.

* **Link to Subsystem Application**: when the subsystem is an online application, the URL address of the subsystem shall be entered.
* **Software Application**: the name of the software application. For example, the software application of the LIS system is Bika system, “Bika System” shall be entered here.
* **Hosting Operating System**: the hosting operating system of the software application, e.g. Windows, Linux.
* **Hosting Type/Source**: the hosting type of the software application, e.g. desktop, server, cloud.

The following fields shall be added manually and be filled for dataset that belongs to the Excel Tools or Resources category.

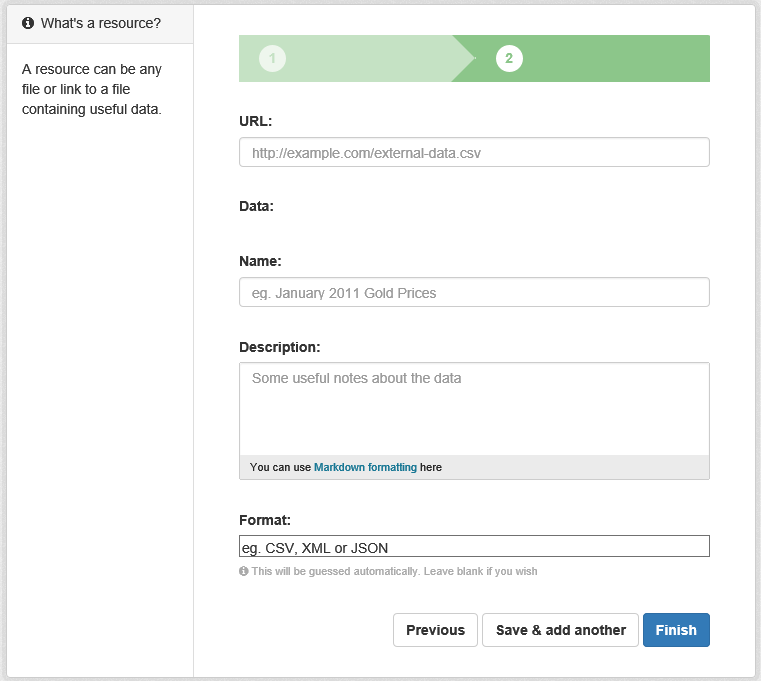
* **Data Location/Link**: if the excel tool or resource can be accessed online, the URL address shall be entered. If the excel tool or resource is stored on locally, the location of the resource shall be provided. If the excel tool or resource is stored both online and locally, both of the information shall be provided.
* **Data Format**: format of the dataset, e.g. PDF, XLS, IMG, paper based form, etc.

If there are other information shall be added for this dataset, user can use the “Custom Field” to add more information.

Note: the fields that are offered by default by the system, but are not mentioned in this guide, do not need to be filled.

#### Add Resources for Dataset

After the necessary information are added for the dataset, user can click the “Next: Add Data” button to add resources for the dataset.



For dataset that belongs to the dataset category, the metadata of the dataset shall be added.

* **URL**: if the metadata is available online, user can enter the URL address that leads to the metadata of the dataset.
* **Data**: by clicking on the blank area below “Data”, a window will be opened where user can upload the metadata from his/her local computer.
* **Name**: the name of the resource shall tell viewers what is the resource about. If all metadata of the dataset is presented in one single file, a recommended naming is “What can you find in the dataset”. If the metadata are separated in multiple files, each file can be name after the content of that piece of metadata.
* **Description**: a short description of what kind of information are included in the metadata shall be provided. If applicable, the meaning of the metadata shall also be briefly described.
* **Format**: the format of the metadata shall be entered here. For example, if the metadata is presented in HTML, “HTML” shall be entered here.

For dataset that belongs to Excel Tools category, two kinds of resources shall be added.

One is an empty excel sheet that gives viewers a brief idea how the tool looks like and what information can they find. The other is the metadata of the Excel tool.

For the empty excel tool:

* **URL**: if the empty excel tool is available online, user can enter the URL address that leads to the empty excel tool.
* **Data**: by clicking on the blank area below “Data”, a window will be opened where user can upload the empty excel tool from his/her local computer.
* **Name**: the name of the resource shall tell viewers what is the resource about. A recommended naming is “Empty Excel Tool”.
* **Description**: a short description of the Excel tool.
* **Format**: the format of the empty excel tool is “XLS”.

For the metadata of the excel sheet:

* **URL**: if the metadata is available online, user can enter the URL address that leads to the metadata of the excel tool.
* **Data**: by clicking on the blank area below “Data”, a window will be opened where user can upload the metadata from his/her local computer.
* **Name**: the name of the resource shall tell viewers what is the resource about. A recommended naming is “What information can you find in the excel tool”.
* **Description**: a short description of the piece of metadata.
* **Format**: the format of the metadata shall be entered here. For example, if the metadata is presented in Excel, “XLS” shall be entered here.

For dataset that belongs to Resources category, if applicable, the actual resources shall be added.

* **URL**: if the resource is available online, user can enter the URL address that leads to the resource.
* **Data**: by clicking on the blank area below “Data”, a window will be opened where user can upload the resource from his/her local computer.
* **Name**: the name of the resource shall tell viewers what is the resource about.
* **Description**: a short description of the resource.
* **Format**: the format of the resource shall be entered here. For example, PDF, DOC, PNG.

If there are other resources to be added for the dataset, user can click on the “Save & add another” button to add more resources.

If there’s no more resources to be added, user can click on the “Finish” button, and the dataset will be created.

### Edit Dataset

To edit a dataset, user can navigate to:

* *Datasets -> select a dataset -> Manage*
* *Organizations -> select an organization -> select a dataset -> Manage*

In the situation that all custom fields are used, one more custom field will be added to the “Edit metadata” tab every time the user opens the edit page of the dataset.

To edit resources within the dataset, user can click on the “Resources” tab.

### View Dataset

To view a dataset, user can navigate to *Datasets -> select a dataset* Or *Organizations -> select an organization -> select a dataset*

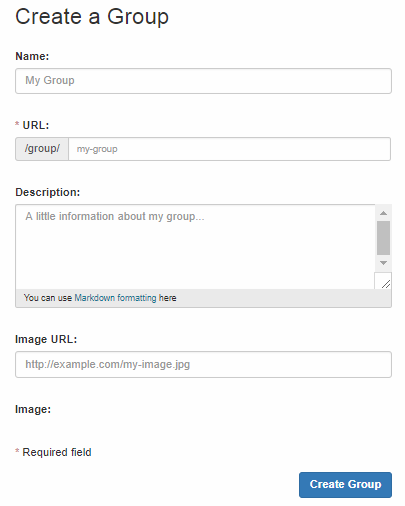
* “Dataset” tab: the “Dataset” tab will be displayed by default. Information about the dataset as well as the resources will be displayed here.
* “Groups” tab: under this tab, user can see to what groups this dataset belong, as well as assign the dataset to other groups.
* “Activity Stream” tab: all updates made to the dataset will be listed under this tab.

## Groups

### Create Group

Groups shall be added according to the categories of datasets. In MoH, the datasets are categories into three categories, Datasets, Excel Tools, and Resources. Therefore, these three groups shall be created.

To create a group, user shall navigate to *Groups -> Add Group*



* **Name**: the name of the group, for example, Excel Tools.
* **URL**: the value entered in this field defines the URL address of this group. The name of the group in lower case shall be entered here with “-” connecting each letter. For example, if the name of this group is “Excel Tools”, the value entered here shall be “excel-tools”.
* **Description**: a brief description of the group shall be entered here. The description shall tell the meaning of this group and what kind of data are included in the group. For example, the explanations of the datasets categories can be entered here.
* **Image URL**: an image/logo of the group can be provided. If the image is available online, user can enter the URL address that leads to the image.
* **Image**: by clicking on the blank area below “Image”, a window will be opened where user can upload the image from his/her local computer.

After clicking on the “Create Group” button, the group will be created. If user wants to add more fields for the group, he/she can do so by clicking on the “Manage” button at the top right corner of the group’s page. Three Custom Fields will appear for user to add more information. This function is the same as for adding “Organizations”.

### Edit Group

To edit the basic information of a group, user can navigate to:

* *Groups -> select* *a group -> Manage -> Edit tab*

Three custom fields will appear at the first time when user tries to edit the group. The same as editing organization, no more custom field will appear when user tries to edit the group again.

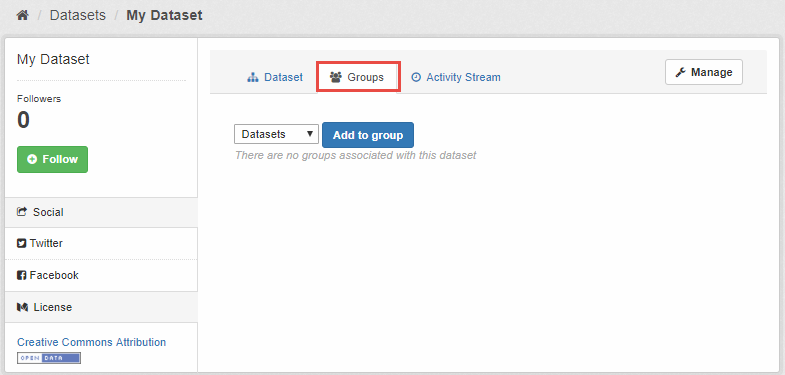
To add members to a group, user can navigate to:

* *Groups -> select* *a group -> Manage -> Members tab -> Add Member*

### Assign Datasets to Groups

To assign a dataset to a group, user shall navigate to the corresponding dataset and then clicks on the “Groups” button.

* *Datasets -> select a dataset -> Groups*
* *Organizations -> select an organization -> select a dataset -> Groups*



User shall choose the group the dataset belongs to, and clicks on the “Add to group” button. The dataset will be assigned to the group. If user clicks on the group, he/she will see that the assigned dataset appears under that group.

Note: It is a good practice to assign every single dataset into a group. By doing so, viewers can easily and accurately see all the datasets, as well as the number of datasets, under each group.

### View Group

To view a group, user can navigate to *Groups -> select a group.*

* “Dataset” tab: the “Datasets” tab will be displayed by default, where user can see all the datasets that belong to the group.
* “Activity Stream” tab: all updates made to the group will be listed under this tab.
* “About” tab: all the additional information about the group added through the Custom Fields will be displayed here.

## User Roles and Access Control Rules

By default, the Data Catalog has the following user roles and access control rules:

* Non-registered user
* can see all organizations and public datasets
* Registered user, not a member of organization
* anyone can self register as a user
* can create organizations, publish datasets within that organization
* the user who created an organization will automatically become the Admin of that organization
* Registered user, organization roles
* member
* can see the organization’s private datasets
* editor
* can edit and publish datasets of that organization
* admin
* can add, remove and change roles for organization members
* System Admin
* created during CKAN setup or later through the server
* customize look and feel of site
* full access to all organizations and datasets
* move datasets between organizations
* permanently delete datasets
* search user by any part of their profile
* delete and edit user’s profile, including username, password

In MoH, the following policy shall be applied:

* Only the System Admin shall be able to create and/or edit organizations. Even though the other types users can do so in the system, he/she shall not create any organizations in the data catalog.
* Only the System Admin shall be able to create and/or edit groups.

## Extensions and References

The data catalog is built on the web-based open source management system CKAN (The Comprehensive Knowledge Archive Network).

The following extensions have been added to the data catalog:

* **Organization Hierarchy**: this extension provides a new field on the organization edit form to select a parent organization. Detailed information about this extension can be found at: <http://extensions.ckan.org/extension/hierarchy/>
* **DataStore**: The CKAN DataStore extension provides an ad hoc database for storage of structured data from CKAN resources. Data can be pulled out of resource files and stored in the DataStore.
  + When a resource is added to the DataStore, you get:
    - Automatic data previews on the resource’s page, using the [Data Explorer extension](http://docs.ckan.org/en/latest/maintaining/data-viewer.html#data-explorer)
    - [The DataStore API](http://docs.ckan.org/en/latest/maintaining/datastore.html#the-datastore-api): search, filter and update the data, without having to download and upload the entire data file
  + The DataStore is generally used alongside the [DataPusher](http://docs.ckan.org/projects/datapusher), which will automatically upload data to the DataStore from suitable files, whether uploaded to CKAN’s FileStore or externally linked.
  + Detailed information about this extension can be found at: <http://docs.ckan.org/en/latest/maintaining/datastore.html>
* **DataPusher**: This application is a service that adds automatic CSV/Excel file loading to CKAN. Detailed information about this extension can be found at: <http://docs.ckan.org/projects/datapusher/en/latest/>
* **text\_view**: Displays files in XML, JSON or plain text based formats with the syntax highlighted. Detailed information about this extension can be found at: <http://docs.ckan.org/en/latest/maintaining/data-viewer.html#available-view-plugins>
* **image\_view**: If the resource format is a common image format like PNG, JPEG or GIF, it adds an <img> tag pointing to the resource URL. Detailed information about this extension can be found at: <http://docs.ckan.org/en/latest/maintaining/data-viewer.html#available-view-plugins>
* **recline\_view**: Adds a rich widget, based on the Recline Javascript library. It allows querying, filtering, graphing and mapping data. Detailed information about this extension can be found at: <http://docs.ckan.org/en/latest/maintaining/data-viewer.html#available-view-plugins>
* **Stats**: CKAN’s stats extension analyzes your CKAN database and displays several tables and graphs with statistics about your site. Detailed information about this extension can be found at: <http://docs.ckan.org/en/latest/maintaining/stats.html>

For more detailed functions of CKAN, user can refer to its documentations: <http://docs.ckan.org/en/ckan-2.7.2/>.