

ACEP Data Catalog User Guides

Jenae Matson

2024-08-05

Table of contents

Welcome	3
1 Getting Started	4
Making an Account	4
Searching for Datasets	5
Submitting a Comment/Complaint	6
2 Researcher	7
Posting a Dataset	7
Tagging a Dataset	9
Editing/Deleting a Dataset	10
Adding a Dataset to a Group	11
Creating a New Group	12
3 Admin	14
Giving User Permissions	14
Reviewing and Publishing a Dataset	15
Deleting Datasets/Groups/Organizations	16
Purge Deleted Datasets/Groups/Organizations	16
Restore Deleted Datasets	17
Restoring a Deleted User	19
For Developers	20
Developing the Data Catalog	20
Basic Docker Commands	20
Creating a Local Instance	21
Create a New Extension	22
Install an Extension	23
Updating the Main Site	23
Extensions	23
Currently Installed	23
Adding Alternate Schemas with ckanext-scheming	25
Attempted Extensions	26

Welcome

This documentation provides some helpful tutorials for working with the ACEP Data Catalog.

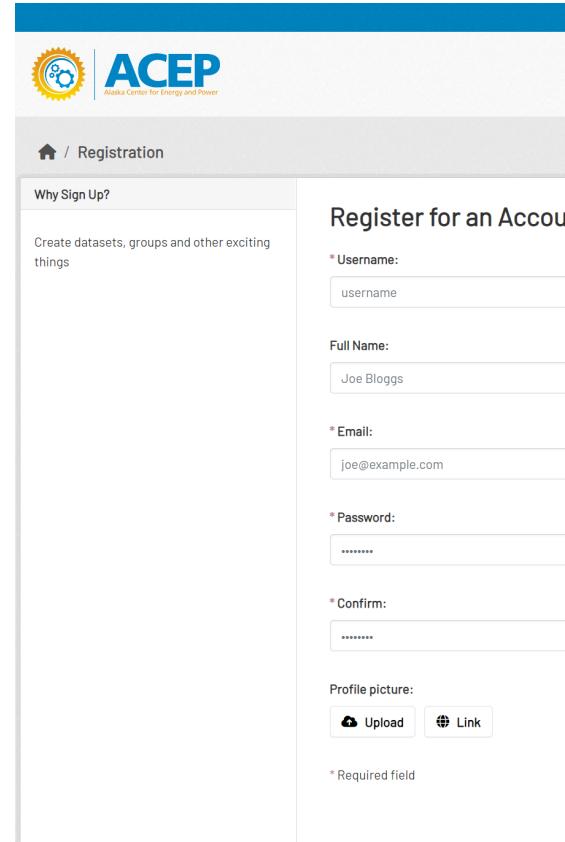


1 Getting Started

Here are some instructions for how to navigate the data catalog.

Making an Account

1. Click the **Register** button in the upper right corner of the screen.



The screenshot shows the ACEP (Alaska Center for Energy and Power) registration page. At the top, there is a logo for ACEP and a subtext "Alaska Center for Energy and Power". Below the logo, the page title is "Registration". On the left side, there is a sidebar with a "Why Sign Up?" section containing the text "Create datasets, groups and other exciting things". On the right side, there is a form titled "Register for an Account". The form fields include:

- "Username": A text input field with the placeholder "username".
- "Full Name": A text input field with the placeholder "Joe Bloggs".
- "Email": A text input field with the placeholder "joe@example.com".
- "Password": A password input field with placeholder text "*****".
- "Confirm": A password confirmation input field with placeholder text "*****".
- "Profile picture": A section with "Upload" and "Link" buttons for profile picture selection.

At the bottom right of the form, there is a note "* Required field".

2. Fill out your information, using your UA email if you have one.
3. In order to see the Internal Use datasets or post datasets, you will need permissions granted to your account. If you are an ACEP employee, contact Liz or another admin.

Searching for Datasets

There are many ways to search for datasets in the catalog.

- **Search Bar**

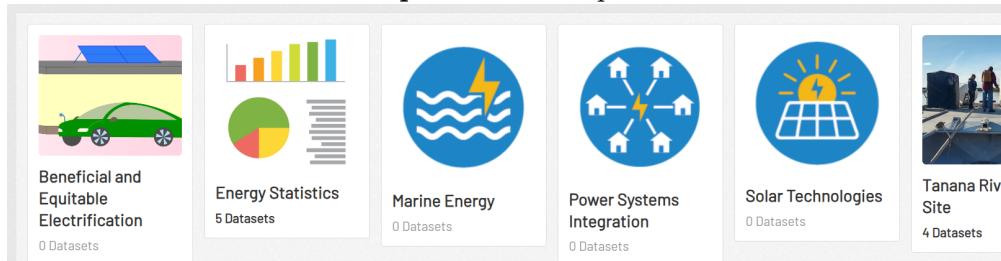
Use key words to find datasets, just like Google. On the left sidebar of the search results page, you can filter by tags and other categories to get more specific results.

The screenshot shows a search results page for the term "Tanana". The left sidebar contains filters for Organizations (ACEP Open Data - 3), Groups (Tanana River Test Site - 3), Tags (Alaska - 3, CEC - 3, Code - 3, Energy - 3, Hydrokinetic - 3, Marine - 3, MATLAB - 3, MHK - 3, Nenana - 3, Power - 3), Formats (none found), and Licenses (none found). The main search bar has "Tanana" entered, and the results are ordered by Relevance. Three datasets are listed:

- Water Horse Hydroelectric Harvester Single Oscillator Field Testing Data,...**
Raw and processed timeseries data generated during field testing of a single oscillating hydrofoil Water Horse prototype at PMEC (Pacific Marine Energy Center) Tanana River Test...
Tags: Alaska, CEC, Code, Current Energy Converter, Energy, Field Test Content Model, Hydrokinetic, MATLAB, MHK, Marine, Nenana, Oscillating Hydrofoil, Power, Pre-Processed Data, Processed Data, Raw Data, River, Tanana River, Technology
- Water Horse Hydroelectric Harvester Dual Oscillator Field Testing Data,...**
Raw and processed timeseries data generated during field testing of a single oscillating hydrofoil Water Horse prototype at PMEC (Pacific Marine Energy Center) Tanana River Test...
Tags: Alaska, CEC, Code, Current Energy Converter, Energy, Field Test Content Model, Hydrokinetic, MATLAB, MHK, Marine, Nenana, Oscillating Hydrofoil, Power, Pre-Processed Data, Processed Data, Raw Data, River, Tanana River, Technology
- Simulink Model of a New Energy EVG-005 5kW Hydropower Turbine**
Simulink model for a New Energy 5kW hydropower turbine. ADCP data ("ds_streamwise_7_13.nc") and DC voltage, DC current, and rotor rotation observed from the New Energy EVG-005...

- **Groups**

Browse through the groups that can be found at the **Groups** tab at the top of the screen



or listed on the home page.

- **Organizations**

Sort by larger categories by looking the **Organization** tab at the top of the screen.

- The **ACEP Open Data** organization holds datasets created and owned by ACEP’s research teams.
- The **ACEP Internal Use** organization holds datasets that ACEP does not own. These are useful resources for ACEP researchers. **You must be an ACEP employee to gain access to these datasets.** Contact Liz to get the necessary permissions for your account.

Submitting a Comment/Complaint

If you find a problem with a dataset or have a comment or suggestion for the data catalog, please contact the ACEP data team at uaf-acep-dcm-support@alaska.edu.

2 Researcher

This is some things researcher's posting stuff should know

Posting a Dataset

There are two ways to access the **Add Dataset** button:

- Click on the **Datasets** tab at the top of the screen
 - Click on the **Organizations** tab at the top of the screen and select the organization you want to add a dataset to:
 - Choose **ACEP Open Data** if you are posting your own data or data that ACEP owns.
 - Choose **ACEP Internal Use** if you are posting a useful dataset that ACEP does not own.
1. Click on the **Add Dataset** button above the search bar.

2. On the first page, fill out the metadata fields for your dataset.
 - Double check the organization field.
 - Choose **ACEP Open Data** if you are posting your own data or data that ACEP owns.
 - Choose **ACEP Internal Use** if you are posting a useful dataset that ACEP does not own.
 - If you are a researcher, your dataset's visibility will automatically be set to private. An admin will review your dataset and make it public.
3. Once you have completed the metadata fields, click the **Next: Add Data** button at the bottom of the form.

4. On the next page, add the data resources to your dataset.
 - Small datasets (<2MB) can be uploaded directly to the catalog
 - Alternatively, enter a link to where the data is stored, such as a GitHub repository or Google Drive.
5. If you have more resources to add, click the **Save & add another** button at the bottom of the form. Otherwise click **Finish** to post your dataset.

Tagging a Dataset

TBD

Editing/Deleting a Dataset

1. Navigate to the dataset page and click on the **Manage** button in the upper right.

The screenshot shows the ACEP dataset management interface. At the top, there's a navigation bar with links for Datasets, Organizations, Groups, About, and FAQ. A search bar is also present. Below the navigation, the dataset title "Alaska Railroad Features" is displayed, along with its status as PRIVATE and a follower count of 0. There are "Follow" and "Organization" buttons. To the right, there's a "Manage" button, which is circled in red. The main content area shows the dataset details: "Alaska Railroad Features" is described as "Geodata displaying various features of the Alaska Railroad". It lists four resources: "Crossings.geojson", "MP.geojson", "Railroad_Boundary.geojson", and "Track_Centerline.geojson", each with an "Explore" button. Below the resources is a tag cloud containing terms like "alaska railroad", "alaska railroad corporation", "mileposts", "railroad", "railroad boundary", "railroad crossings", and "track centerline". A "Metadata" section is also visible at the bottom.

2. This displays the **Edit metadata** page where you can change the metadata of the dataset. After making changes, click the **Update Dataset** button at the bottom of the form.

- To delete the dataset, click the **Delete** button at the bottom of the form.

NOTE: Deleting a dataset does not remove it completely from the database. The url of the deleted dataset will not be able to be reused until it has been purged by a sysadmin.

3. To edit the resources or add more resources, click on the **Resources** tab.

- To delete a resource, select it from the list and then click the **Delete** button at the bottom of the page.

Adding a Dataset to a Group

1. Navigate to the dataset page and click on the **Groups** tab.
2. Select a group from the dropdown menu and click the **Add to group** button.
3. To remove the dataset from a group, hover your cursor over the group and click the **Re-**

The screenshot shows the ACEP internal use interface. At the top, there's a navigation bar with links for Datasets, Organizations, Groups, About, and FAQ, along with a search bar and user profile icons. Below the navigation is the ACEP logo and the text "Alaska Center for Energy and Power". The main content area shows the organization "Alaska Railroad Features". On the left, there's a sidebar with a "Follow" button and a "Power Systems Integration" section containing a lock icon. The main panel has tabs for "Dataset" and "Groups". Under the "Groups" tab, there's a list with one item: "Power Systems Integration" (with a red circle around it) and a "Remove" button.

move button.

Creating a New Group

Groups are a good way to group together datasets that are connected. You may want to create groups for common research themes, funding organizations, or projects. 1. Click on the **Groups** tab at the top of the screen. 2. Click the **Add Group** button.

The screenshot shows a web application interface for creating a group. At the top, there's a blue header bar with the ACEP logo and navigation links for Datasets, Organizations, Groups (which is highlighted in blue), About, and FAQ. To the right of the header is a search bar with a magnifying glass icon. Below the header, the URL 'portal.lab.acep.uaf.edu/group/<group>' is visible. The main content area has a light gray background. On the left, there's a sidebar with a section titled 'What are Groups?' containing text about using CKAN Groups to manage datasets. The main form is titled 'Create a Group'. It includes fields for 'Name' (with 'My Group' entered), 'Description' (with placeholder text 'A little information about my group...'), and 'Image' (with two options: 'Upload' and 'Link'). A note at the bottom of the form says '* Required field'. At the bottom right of the form is a blue 'Create Group' button.

3. Enter the information for the group. Find a logo or simple image to upload to represent the group.
4. Once you have entered the information, click the **Create Group** button at the bottom of the form.

3 Admin

Here is some stuff admins should know how to do.

Giving User Permissions

Users have roles in each organization that give them different permissions.

The roles are:

- **Member:** can see private datasets in the organization
- **Editor:** can add private datasets to the organization and delete datasets from the organization
- **Admin:** can change users' roles in an organization and publish datasets from private to public

To give a user a role in an organization:

1. Click on the **Organizations** tab at the top of the screen and click on the organization you want to add the user to.
2. Click on the **Manage** button in the upper right.

The screenshot shows the ACEP organization management interface. At the top, there is a navigation bar with tabs for Datasets, Organizations (which is highlighted in blue), Groups, and About. Below the navigation bar, there is a large icon of a house with a padlock inside it, labeled "ACEP Internal Use" and "ACEP researchers use data from other". To the right of the icon, there is a section titled "9 members" with a table showing three users: "jimatson" (Admin), "test_editor" (Editor), and "test_admin" (Admin). Above the table, there are buttons for "Edit", "Datasets", and "Members" (which is circled in red). There is also a "Add Member" button.

User	Role
jimatson	Admin
test_editor	Editor
test_admin	Admin

3. Click on the **Members** tab at the top.

- To add a new member to the organization, click the **Add Member** button. Enter their username, select their new role from the dropdown, and click the **Add Member** button at the bottom of the form.

Add Member

Existing User : If you wish to add an existing user, search for their username below.

Username:

New User : If you wish to add a new user, enter their email address.

Email address:

Role:

Add Member

- To change an existing member's role, click the wrench next to their user name.

test_editor	Editor	
-------------	--------	--

- Select their new role from the dropdown menu and click **Update Member**
- To remove the user from the organization completely, click the **Delete** button.

Edit Member

test_editor

Role:

Delete **Update Member**

Reviewing and Publishing a Dataset

1. Navigate to the dataset you would like to publish.
2. Review the metadata and ensure the files/links are correct and function properly.
3. To publish, click on the **Manage** button in the top right of the dataset page.

4. In the metadata fields, find the **Visibility** field.

The screenshot shows a form for editing a dataset. At the top, there is a dropdown menu labeled "Organization" with "ACEP Open Data" selected. Below it is a section labeled "Visibility" with three options: "Private" (selected), "Private" (highlighted in blue), and "Public". A "Source" field contains the URL "https://arcg.is/4j5b50".

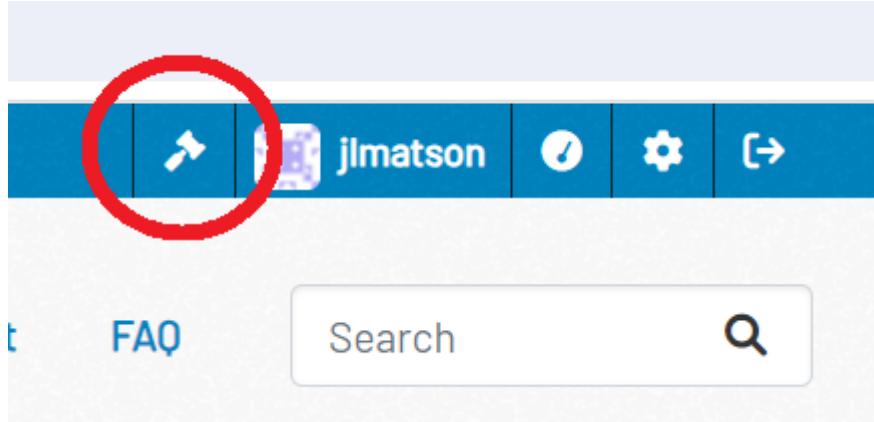
5. Change the **Visibility** field from **Private** to **Public**, and click the **Update Dataset** button at the bottom of the form.

Deleting Datasets/Groups/Organizations

1. To delete a dataset, group, or organization, navigate to its page and click the **Manage** button in the upper right.
2. Click the **Delete** button at the bottom of the form to delete.
3. **Deleting a dataset, group, or organization does not remove it from the database.** After deleting, a sysadmin can restore the item or purge it to remove it from the database. If not purged, the url for the deleted item cannot be reused.

Purge Deleted Datasets/Groups/Organizations

1. Navigate to sysadmin settings by clicking the hammer icon at the top of the page.



2. Click on the **Trash** tab. This page will list all the currently deleted datasets, groups, and

The screenshot shows the ACEP web application interface. At the top, there is a navigation bar with links for Datasets, Organizations, Groups, About, FAQ, and a search bar. Below the navigation bar, a user profile for 'jimatson' is displayed. The main content area is titled 'Trash' and contains instructions: 'Purge deleted datasets, organizations or groups forever and irreversibly.' A red circle highlights the 'Trash' tab in the top navigation bar. On the right side of the trash page, there are three sections: 'Deleted datasets' (listing 'Avalanche Zones' with a 'Purge' button), 'Deleted organizations' (listing none), and 'Deleted groups' (listing 'ARCTIC Program' with a 'Purge' button). A 'Purge all' button is located at the top right of the trash page.

organizations.

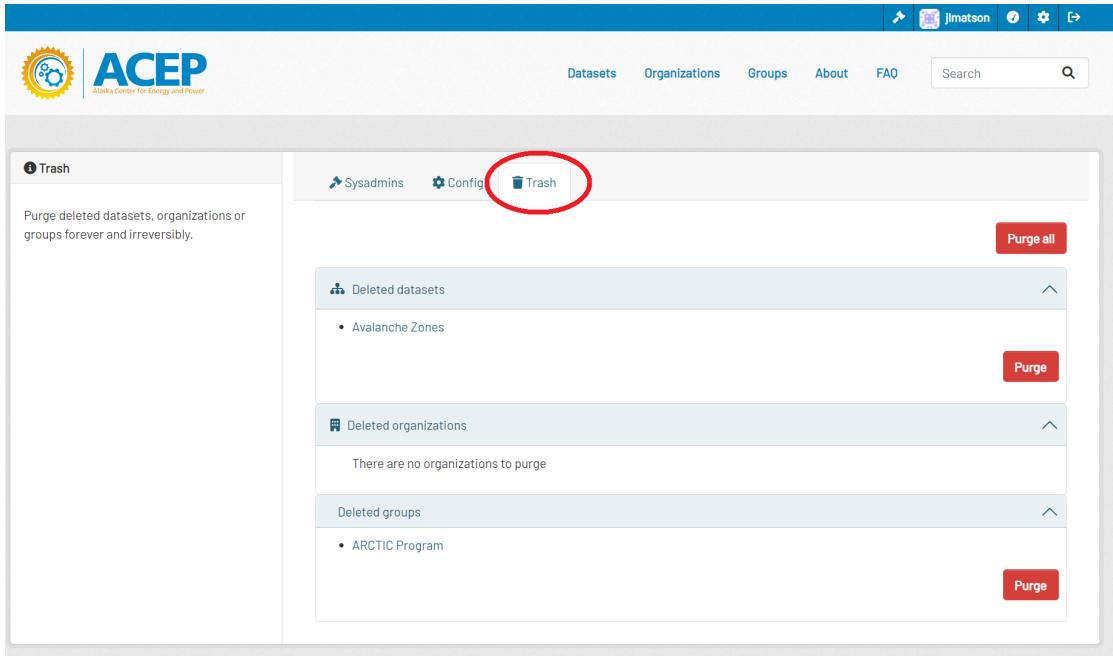
3. You can purge all deleted items using the **Purge all** button at the top of the form. To purge only one group of items (datasets, group, or organizations) use the **Purge** button in that section.

Restore Deleted Datasets

1. Navigate to sysadmin settings by clicking the hammer icon at the top of the page.

The screenshot shows the ACEP web application interface. At the top, there is a navigation bar with links for Datasets, Organizations, Groups, About, FAQ, and a search bar. Below the navigation bar, a user profile for 'jimatson' is displayed. A red circle highlights the hammer icon in the top navigation bar, which is part of the sysadmin settings. On the left side of the page, there are links for 't' (likely 'Logout'), 'FAQ', and a search bar. On the right side, there are icons for a gear (Settings), a lock (Security), and a right-pointing arrow (Logout).

2. Click on the **Trash** tab. This page will list all the currently deleted datasets, groups, and organizations.



3. Click on the dataset you would like to restore, and click the **Manage** button in the upper right.

Visibility:

Private

State:

Deleted

Source:

<http://example.com/dataset.json>

A red circle highlights the "Deleted" option under the State field.

4. In the metadata fields find the **State** field.
5. Change the **State** field from **Deleted** to **Active**, and then click the **Update Dataset** button at the bottom of the form.

Restoring a Deleted User

When a user is deleted from the website, their information remains in the database with the `state` field set to `deleted`. To reactivate the user, you must set this field to `active` in the database.

1. Enter the `acep-db-cont` docker container
 - `> docker exec -it acep-db-cont /bin/bash`
2. Access the postgres database
 - `> psql -U postgres`
3. List all the databases and connect to the `ckandb` database.
 - `> \l`
 - `> \c ckandb`
4. List all the tables and list the columns of the user table.
 - `> \d`
 - `> \d user`
5. List all the users in the user table.
 - `> SELECT * FROM public.user`
6. Find the deleted user with the username [username].
 - `> SELECT id, name, email, state FROM public.user WHERE name = '[username]'`
7. Update the user's state field.
 - `> UPDATE public.user SET state = 'active' WHERE name = '[username]'`
8. Find the user again and ensure that the state field is set to active.
 - `> SELECT id, name, email, state FROM public.user WHERE name = '[username]'`

For Developers

Documentation of the development of the ACEP Data Catalog.

For more information and guides, visit the official [CKAN Documentation](#)

Developing the Data Catalog

The ACEP Data Catalog is run on a VM hosted by RCS. Extensions can be updated by pushing to the aceportal-ckan GitHub repository. After pushing, changes take ~30 min to update on the main site.

Basic Docker Commands

List all running containers:

```
docker ps -a There are 5 containers that run the data catalog
```

- `acep-ckan-cont`
- `acep-db-cont`
- `acep-redis-cont`
- `acep-solr-cont`
- `acep-datapusher-cont`

Rebuild and spin up containers:

```
docker compose up -d --build
```

Go into container:

```
docker exec -it [container_name] /bin/bash
```

Or if bash is not installed in the container:

```
docker exec -it [container_name] /bin/sh
```

Or if in a bash terminal:

```
docker exec -it [container_name] bash
```

Restart a container:

```
docker restart [container_name]
```

Creating a Local Instance

Creating a local version of the data catalog is a useful tool for developing and testing new features.

1. Install Docker: <https://www.docker.com/get-started/>
2. Clone the ACEP CKAN repository from Github: <https://github.com/UAF-RCS/acepportal-ckan.git>
3. Create the `.env` file inside the main `acepportal-ckan` folder. Copy the contents from the `.env.example` file.
4. Specify the location of the source files, storage files, backups, etc. in the `.env` file. You will move those files to these locations in the next steps. For example:

```
# CKAN Mounts Directory
CKAN_EXTENSIONS_MOUNT=./ckan-extension
SRC_EXTENSIONS_PATH=/srv/app/src_extensions
CKAN_SOURCE_MOUNT=./ckan-src/src
CKAN_STORAGE_MOUNT=./ckan-src/storage
CKAN_INI_MOUNT=./ckan-src/ckan.ini
```

5. To create a replica of the current main Data Catalog, copy over the source files, storage files, `ckan.ini` file, and database backups from the VM. These files are located on the VM inside `/opt/ckan/backups`. Use `scp` to copy the files onto your machine. These backups are created everyday: replace [date] with the most recent date in the format `yyyymmdd`.

Inside of `acepportal-ckan/ckan-src` run the following

- `scp user@portal.lab.acep.uaf.edu:/opt/ckan/backups/app_[date].tar.bz2`
- `scp user@portal.lab.acep.uaf.edu:/opt/ckan/backups/app_storage_[date].tar.bz2`
- `scp user@portal.lab.acep.uaf.edu:/opt/ckan/acepportal-ckan/ckan-src/ckan.ini`

6. Use tar to decompress the source and storage tar files

- `tar -jxvf app_[date].tar.bz2`
- `tar -jxvf app_storage_[date].tar.bz2`

Decompressing the `app_storage` tar file should create a folder called `ckan` containing the folders `resources`, `storage`, and `webassets`. Rename the `ckan` folder to `storage`. This should result in the directory structure specified in `ckan-src/README.txt`

7. Create a backups folder alongside the acepportal-ckan repository on your machine. Specify the name in the BACKUP_TO setting in the .env file.

```
# Backups  
BACKUP_TO=.../.../[backups folder name]
```

8. Run the following commands inside the backups folder to copy over the database and datastore.

- scp user@portal.lab.acep.uaf.edu:/opt/ckan/backups/ckandb_[date].tar .
- scp user@portal.lab.acep.uaf.edu:/opt/ckan/backups/datastore_[date].tar .

9. Inside of the ckan.ini file, set the ckan.site_url setting to the localhost url as so:

```
ckan.site_url = http://127.0.0.1:5000
```

10. Build the containers using,

- docker compose up

11. Once the containers are up, use the import_database.sh bash script to import the database.

- bash import_database.sh

12. Rebuild the CKAN search index.

- docker exec -it acep-ckan-cont /bin/bash
- cd /srv/app
- ckan search-index rebuild

Create a New Extension

1. Enter the acep-ckan-cont Docker container

- docker exec -it acep-ckan-cont /bin/bash and run the following command
- ckan generate extension -o /srv/app/src/ckan-extension This will create an extension in the ckan-extension folder which can be edited outside of the container.

2. Add the extension name to the CKAN_PLUGINS list in the .env file.

3. Run docker compose up -d --build ckan

Install an Extension

1. Ensure that the extension supports CKAN 2.10.4 and Python 3.10. Clone the extension repository into the `ckan-extension` folder.
2. Ensure that all dependencies for the extension are listed in `requirements.txt` or a similar file.
3. Add the extension name to the `CKAN_PLUGINS` list in the `.env` file.
4. Run `docker compose up -d --build ckan`

Updating the Main Site

To add a feature from your local instance to the main Data Catalog,

1. Push the files to the `aceportal-ckan` GitHub repository.
2. Wait about 30 min. for the changes to be pulled to VM.
3. If you have added a new extension, ssh into the VM and add the extension name to the `.env` file.
- 4.

Extensions

Currently Installed

ckanext-customtheme

Author: Jenae Matson

Purpose: Add custom theming and features for the CKAN instance, including

- ACEP logos, colors, and fonts
- Home page layout, images, and featured dataset
- Changed font weight of Register button
- Added tags to search page display
- HTML file for About page text
- Removed social media links from dataset/resources pages
- Added support contact info to dataset sidebar
- Added default blank option to add-to-group dropdown menu

ckanext-faqpage

Author: Jenae Matson

Purpose: Create an FAQ page linked in the masthead with collapsible boxes for questions and answers.

ckanext-restrictpublish

Author: Jenae Matson

Purpose: Restrict the ability to change the visibility of a dataset to admins only. Datasets posted by editors default to private.

ckanext-geoview

Link: <https://github.com/ckan/ckanext-geoview>

Purpose: Created resource views for geojson and other geo-data file types. We have implemented the OpenLayers Viewer.

ckanext-package-group-permissions

Link: <https://github.com/salsadigitalauorg/ckanext-package-group-permissions>

Purpose: Allows all editors and admins to add datasets to any group, without having to be added as members to each group.

Modifications: This extension was created and works with CKAN 2.9. This instance is version 2.10, so the extension requires some small modifications to work. The following changes were made to the original extension:

- In the file `plugin.py`, change the `member_create` function to the following

```
def member_create(self, next_auth, context, data_dict):  
    """  
    This code is largely borrowed from /src/ckan/ckan/logic/auth/create.py  
    With a modification to allow users to add datasets to any group  
    :param context:  
    :param data_dict:  
    :return:  
    """  
  
    group = logic_auth.get_group_object(context, data_dict)  
  
    authorized = False  
    if not group.is_organization and data_dict.get('object_type') == 'package':
```

```

authorized = helpers.user_has_admin_access(include_editor_access=True)

if not authorized:
    # Fallback to the default CKAN behaviour
    return next_auth(context, data_dict)
else:
    return {'success': True}

```

- In the file `templates/package/group_list.html`, add the line `{ h.csrf_input() }` to the beginning of the two post forms, as follows

```

{% if groups %}
<form class="add-to-group" method="post">
    {{ h.csrf_input() }}
    ...
</form>
{% endif %}

```

```

{% if c.pkg_dict.groups %}
<form method="post">
    {{ h.csrf_input() }}
    ...
{% endif %}

```

ckanext-scheming

Link: <https://github.com/ckan/ckanext-scheming>

Purpose: Allows for the creation of alternate metadata templates (schemas) defined by .yaml or .json files.

Adding Alternate Schemas with ckanext-scheming

1. Create a .yaml or .json file in the folder `ckanext-scheming/ckanext/scheming` to define the metadata schema. See extension documentation for more information and examples.
2. In `ckan.ini`, add your schema to the `scheming.dataset_schemas` config option. For example:

```
scheming.dataset_schemas = ckanext.scheming:arctic_dataset.json  
                           ckanext.scheming:geo_dataset.json
```

3. The new dataset creation form is located at a url defined by the schema type name. For example, the creation form for datasets of type `arctic-dataset` is located at `/arctic-dataset/new`. You can define a new Add Dataset button using this new url.

Attempted Extensions

ckanext-spatial

Link: <https://github.com/ckan/ckanext-spatial>

Purpose: This extension adds the ability to search for datasets on a map widget, as well as a dataset extent map widget on the dataset page, provided correct geospatial metadata.

Problems: This extension is not currently installed due to the following,

- Configuring map tiles for ckanext-spatial caused the map tiles for ckanext-geoview to disappear.
- Datasets with the required spatial metadata were not searchable on the map search widget, although the dataset extent widet worked correctly.

ckanext-oidc-pkce

Link: <https://github.com/DataShades/ckanext-oidc-pkce/tree/master>

Purpose: This extension allows for users to be authenticated through an external application when they login.

Problems: Ideally users on the ACEP Data Catalog would be able to login using their UA login credentials through Google Authentication. This extension installs correctly, but does not seem to support Google Authentication.