Data Organization in Clowder

Table of Contents

About this Page

This page contains some initial ideas about how to organize data within Clowder for this project.

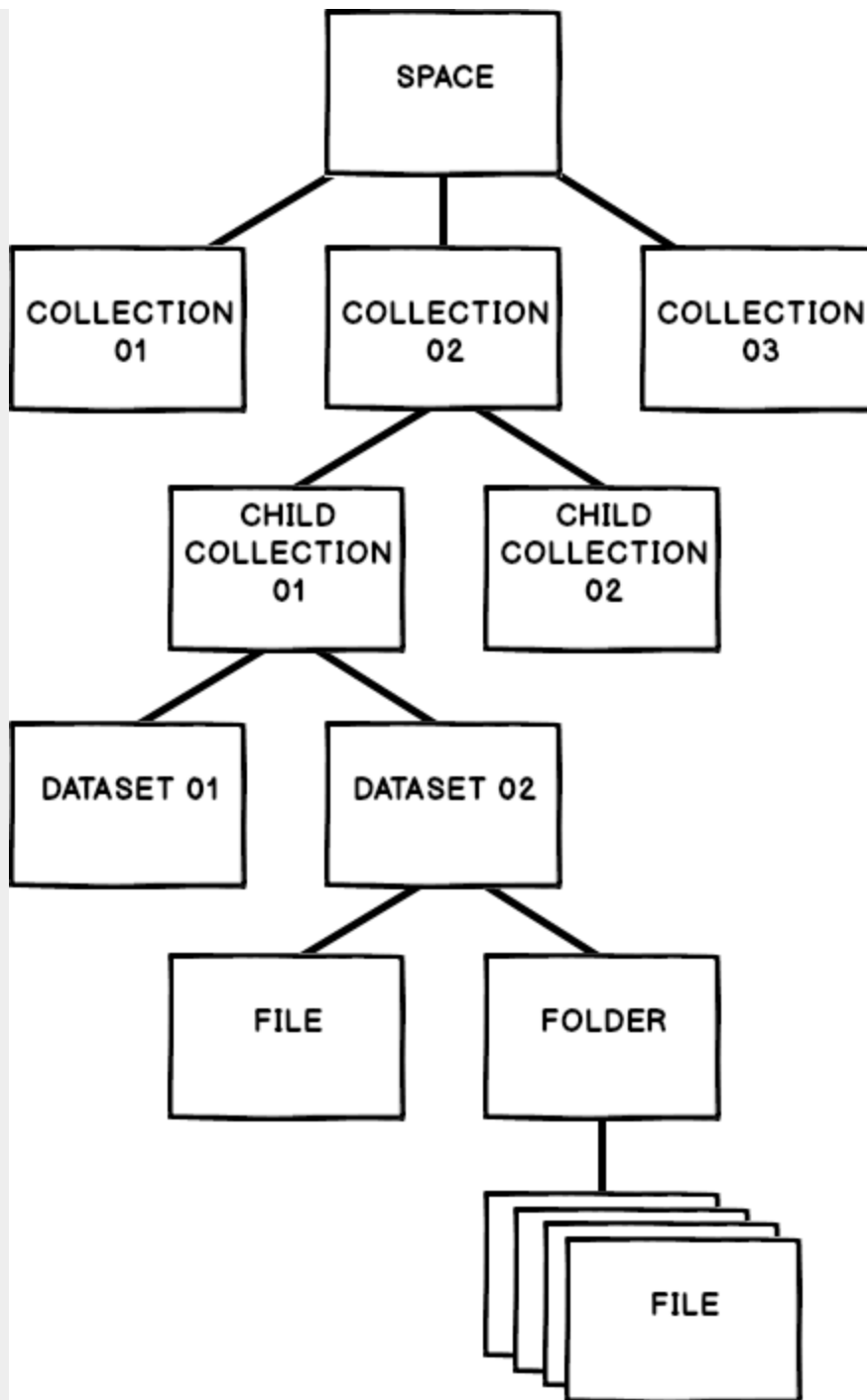## Clowder Information

General

There are three basic sections within Clowder:

- Spaces
- Collections
- Datasets

Access Control Details:

- Being a Clowder User does not automatically provide Access to all Spaces
- Access type is established at the Space level (Admin, Editor, Viewer)
- Access must be set for each Space individually
- All Datasets and Collections inherit permissions from the Space to which they belong

Graphical Description

SPACE

COLLECTION 01

COLLECTION 02

COLLECTION 03

CHILD COLLECTION 01

CHILD COLLECTION 02

DATASET 01

DATASET 02

FILE

FOLDER

FILE

Organization Ideas

Naming Conventions

Some ideas for how to name items that are placed in Clowder:

- When able, include the name of the Sensor in the title of Spaces, Collections, and Datasets
- Including the name of the Sensor may make it is easier to search for items

Metadata and Tags

Metadata and Tags can be utilized to enhance organization and search capabilities.

- Metadata
    - Metadata is *searchable*
    - Datasets can have Metadata
    - Files (in either Datasets or Folders) can have Metadata
    - Definition sources are provided for each Metadata field
- Tags
    - Tags are *searchable*
    - Can view a list of all Tags in the application
    - Datasets can have Tags
    - Files (in either Datasets or Folders) can have Tags

Using the Sections Within Clowder

It would be good to have a general organization methodology moving forward.

This section contains ideas for organization.

An idea for utilizing and organizing these sections:

- Spaces
    - Named after each Sensor or a general category (e.g.: "How To" Documentation, Images, etc.)
    - Contains both Collections of Datasets and individual Datasets
    - Can have an image appear in the listing on Clowder
- Collections
    - Naming would include the Sensor name
    - Contains Datasets
- Datasets
    - The name would include the name of the Sensor
    - Should be placed in Collections
    - Some Datasets can contain raw data and other derived data
    - Can contain Folders that contain files
    - Datasets and Files can both have Metadata and Tags

This is an example illustrating this idea:

- This is based upon the Flux Tower data (information about this Dataset is available in this Wiki space at **Flux Tower Data**)
- **Space**
    - There is only one of these for this Space
    - *Space Name:* Flux Tower Site
    - *Space Description:* Flux Tower located at `location-description`
    - *Space Image:* Suggestion - provide an image of the site or the instrument or a logo (if available)
- **Collection**
    - There can be several of these for this Space
    - Raw Data Measurements Example
        - *Collection Name:* Flux Tower Raw Data Files
        - *Collection Description:* Collection of Raw Data Files
        - *Collection Content:* This will contain all relevant Datasets for this collection
    - Data Measurements Example
        - *Collection Name:* Flux Tower Data Files
        - *Collection Description:* Collection of Data Files
        - *Collection Content:* This will contain all relevant Datasets for this collection
    - Documentation Example
        - *Collection Name:* Flux Tower Documentation
        - *Collection Description:* Documentation for the Flux Tower Sensor
        - *Collection Content:* This will contain any relevant documentation that describes the Sensor (types of measurements gathered, etc.)
            - Datasets would be present in this collection for enhanced organization
            - Files can be of any preferred format
- **Datasets**
    - There can be several of these for each Collection in the Space
    - To simplify managing data, use Folders in yearly Datasets to group files by month
    - Datasets with Folders Example
        - Would be in the Flux Tower Raw Data Files Collection
        - *Dataset Name:* Flux Tower Raw Data 2016
            - Flux Tower `information` for 2016
            - The `information` would just describe the general content of the Dataset (type of measurement, type of data, etc.)
        - *Dataset Description:* Flux Tower Raw Data collected during the year of 2016
            - A Dataset with Flux Tower `information` for 2016
        - *Dataset Metadata:* This will contain any relevant Metadata
        - *Dataset Tags:* This will contain any relevant Tags
        - *Dataset Content:* This will contain Folders for each month of data
            - *Folder 1:* January 2016
                - This will contain all January files
            - *Folder 2:* February 2016
                - This will contain all February files
            - *Folder 3:* March 2016
                - This will contain all March files
        - Keep creating Folders for the year as needed

Organization Tasks

General Tasks to be completed:

- Decide upon the organization methodology
- Removing any extraneous or duplicated data
    - Remove "imlczo" Space and move Datasets contained within to appropriate Spaces
- Where possible, moving and renaming items to fit the decided upon organization methodology
- Provide a "**How To**" wiki page about the decided upon organization methodology
- Make some example changes in Clowder that reflect the ideas presented herein