

Extractor Catalog

- [Overview](#)
- [Order and Options \(independent of priorities\)](#)
- [Suggested Improvements](#)
 - [High-Level Requirements](#)
 - [Specific Action Items](#)
 - [Navigation and Privacy](#)
 - [Discovery and Search](#)
 - [Organization](#)
 - [Diagnostics, Maintenance, and Feedback](#)
 - [Changing Assumptions](#)
- [Mockups](#)
 - [Catalog View](#)
 - [Log Viewer](#)
 - [Rate & Comment](#)
 - [Extractor Details View](#)
 - [Label Management View](#)
 - [Create New Label](#)
 - [Assign Labels](#)
 - [Comments View](#)
 - [Rate & Comment](#)
 - [History & Metrics View](#)

Overview

Clowder currently offers a list of extractors that have been registered to the system. When there are too many extractors registered in the system, it can be overwhelming to the user and make it difficult for them to find the extractor they need. Furthermore, a lack of categorization makes it difficult to discover new extractors or to suggest improvements to existing extractors' maintainer.

Current behavior does not allow for an extractor to be "hidden", as the current admin-only list contains all extractors registered in Clowder.

Order and Options (independent of priorities)

First, we need to decide if this Extractor Catalog should be separate from the current admin-facing UI for listing extractors. The main question being: is this new Extractor Catalog a) an evolution of our current admin view, or b) a new view borne from new objectives with a new focus. There are pros and cons to each approach.

Option A yields a head-start on some of the boilerplate work of setting up the view, but adds the additional work of locking that view down to make sure it is truly read-only for normal users while maintaining existing admin functionality.

Option B yields a more user-focused UI that should prove to be more easily testable than one that also contains all of the admin functionality, at the cost of some additional boilerplate work in setting up the new view.

Suggested Improvements

We are proposing an enhanced view or set of views for searching and discovering extractors within the catalog.

High-Level Requirements

First and foremost, we will need to expose this catalog to non-admin users.

Some organizational tools would be very helpful, such as grouping extractors into "toolboxes" with a simple string tag as an identifier. This identifier could be indicative of the use or function for the extractor, or could indicate which group uses the extractor, or could even be completely arbitrary. Administrators should be allowed to choose who can Create new toolboxes for further classification of extractors.

The catalog should be sortable by extractors used the most by the user viewing the list.

Furthermore, the catalog should be searchable/filterable by one or more of the following:

- a particular space
- a particular file trigger
- a particular metadata term

It would also be nice to have some additional methods to debug or collect information about an extractor. There should be a way to view the extractor logs, as well as a way to be notified when a new version of an extractor is released. Job history and metrics should be made available, along with a flag indicating whether this extractor is ready for public consumption.

Specific Action Items

Navigation and Privacy

- View catalog from within Clowder - decide if this is a new view or improving an old view
- Expose extractor catalog to authenticated users - hide admin-only function behind permissions in the UI
- Navigation to catalog entry by clicking on name - self-explanatory, should be easy once we have a catalog view to link to
- Limit who can create toolboxes - widget or view for listing users and setting permissions

Discovery and Search

- View all extractors maintained by my project - given a project/space, list extractors that are being used in that space
- See projects/spaces using a particular extractor - given an extractor, list projects/spaces
- Search for extractors that are visible to the entire site - there is currently no "search" for extractors, so this may constitute some new API calls
- Discover extractors that are private to a space - link from a space to extractor catalog with search prepopulated?
- Search by file triggers for a specific extractor - widget+API for providing arbitrary file trigger patterns
- Search by metadata term - widget+API for providing specific metadata term and/or value

Organization

- Core set of metadata for Clowder to operate - allow admins to tag extractors as they currently can with files/datasets/collections
- Core set of metadata on our instance's catalog - allow admins to tag extractors as they currently can with files/datasets/collections, inherit from above (indicate sources of tags, if possible)
- Core set of metadata for specific toolbox - allow admins to tag extractors as they currently can with files/datasets/collections, inherit from above 2 (indicate sources of tags, if possible)
- Sort my extractors by usage - list most recently/heavily used extractors first

Diagnostics, Maintenance, and Feedback

- See the logs files for an extractor - self-explanatory, linking to graylog or similar may be enough
- Users should be able to comment on and rate extractors - new widget+API for comments and ratings, new db collection(s)
- I want to know the job history of an extractor within the catalog - can lean on existing APIs where possible
- Stats on extractor page based on metrics - we are not yet collecting metrics such as CPU/MEM usage, so we will need to talk about how that will work, but we can show # of uses, top users, etc
- Notifications with extractor version changes - how do we let users know that a newer version of an extractor has been deployed?
- Flag extractor in dev, staging, prod - allow extractor maintainers to set a "development status" on their running extractor to let users know if it is ready to use

Changing Assumptions

Some assumptions within Clowder will be changed by carrying out the above directives, including but not limited to:

- Extractor list is no longer admin-only
- Extractor list is no longer read-only
- Extractors are no longer public to entire instance, can now be "private to a space"
- With the addition of multiple versions, extractor name may no longer work as the primary key in the database depending on our implementation (arguably, this should change either way)

Mockups

For presentation to the user, I propose using the term "label" instead of "toolbox" or "tag" or "group".

This term is not currently overloaded within Clowder and in my opinion more accurately conveys that this is simply a string identifier used for organization.

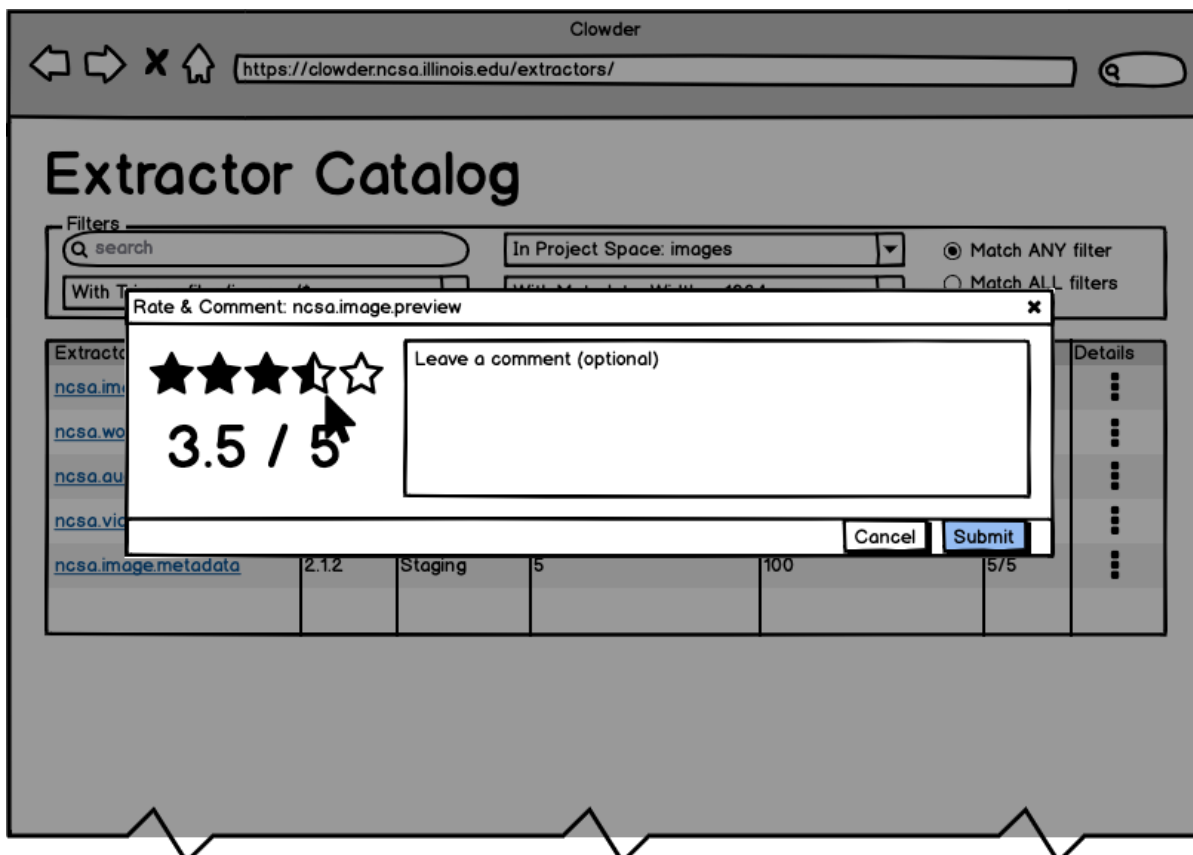
Catalog View



Log Viewer



Rate & Comment



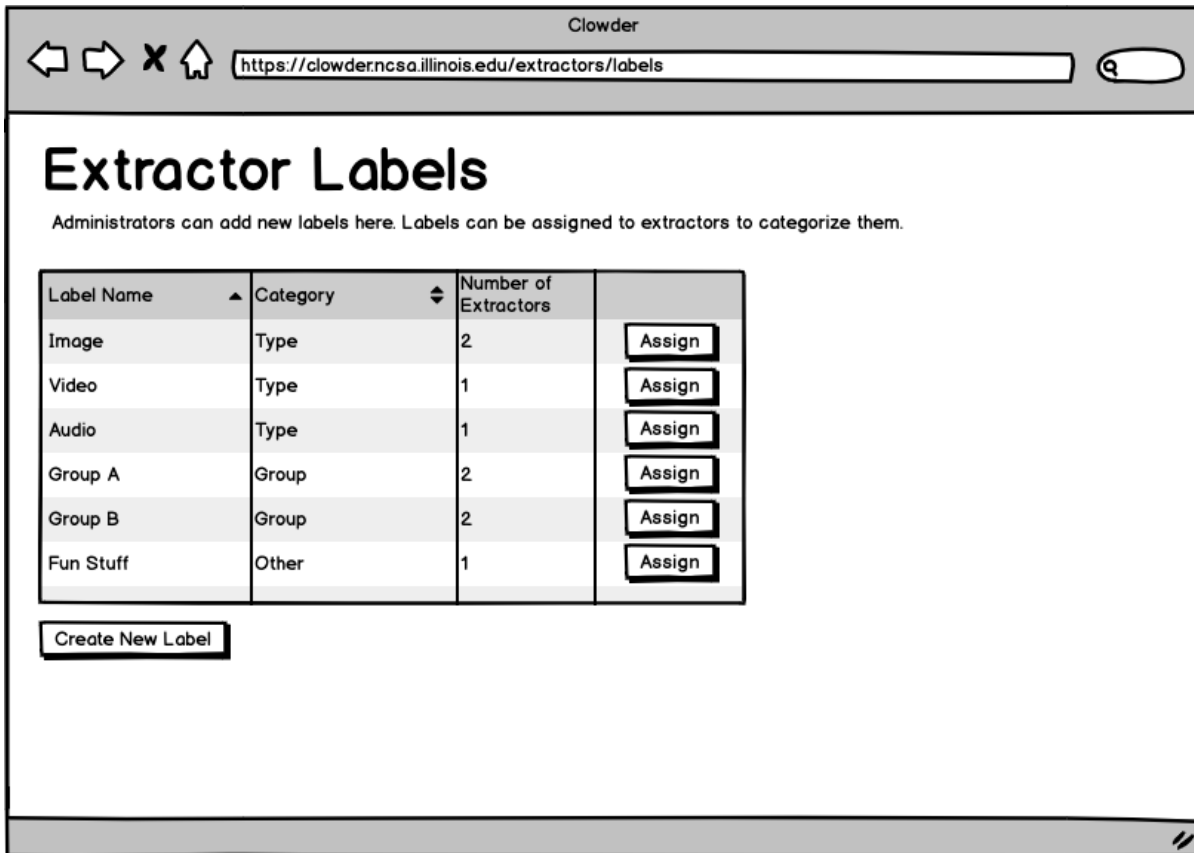
Extractor Details View

The screenshot shows a web browser window with the Clowder logo in the top right. The address bar displays the URL `https://clowder.ncsa.illinois.edu/extractors/ncsa.image.preview`. The main content area is titled **ncsa.image.preview** and contains the following sections:

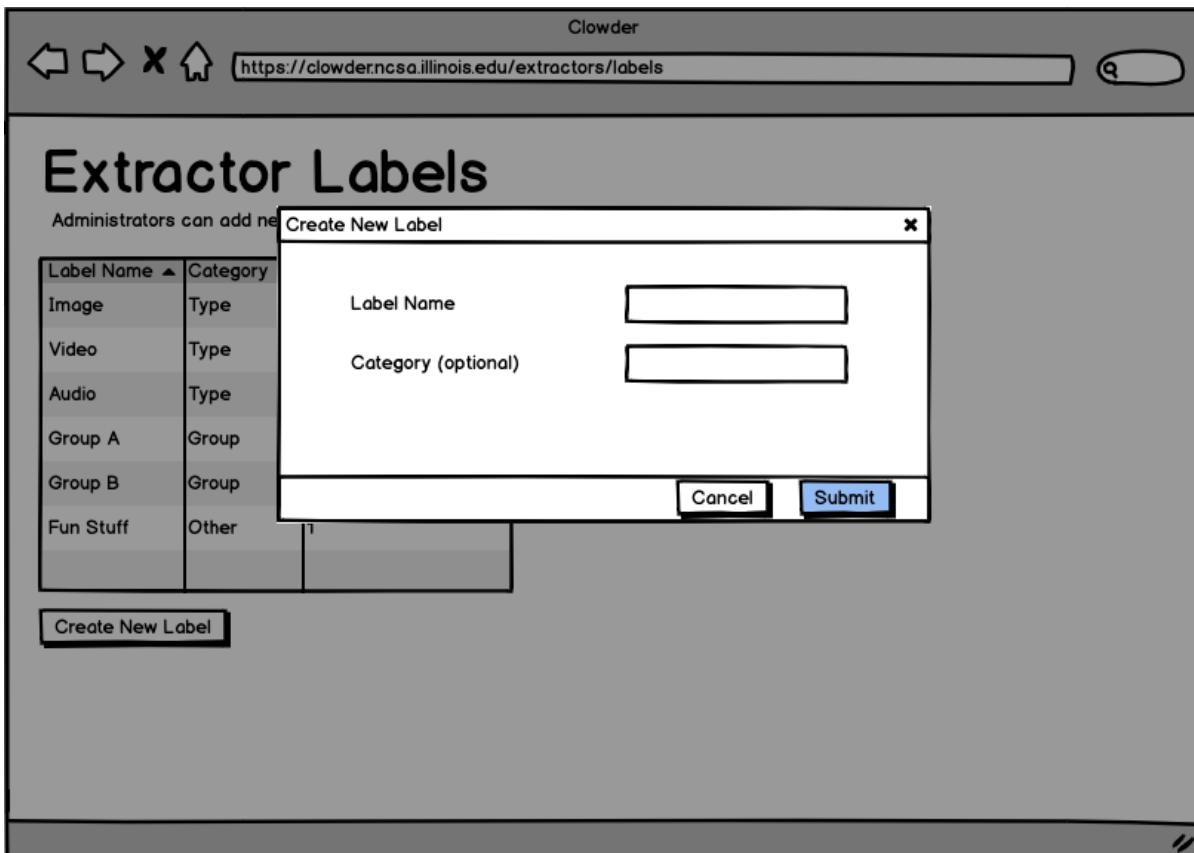
- Description:** Creates thumbnail and image previews of Image files.
- Author:** [Rob Kooper <kooper@illinois.edu>](#)
- Version:** 1.0
- Repositories:**
 - Git:** <https://opensource.ncsa.illinois.edu/bitbucket/scm/cats/extractors-core.git>
 - Docker:** [clowder/extractors-image-preview](#)
- Contributors:**
 - [Sandeep Putharveetil Satheesan <sandeeps@illinois.edu>](#)
- File Triggers:**
 - image/*
- Labels:**
 - Image
 - Group A
 - Fun Stuff

A **Manage** button with a pencil icon is located to the right of the Labels section.

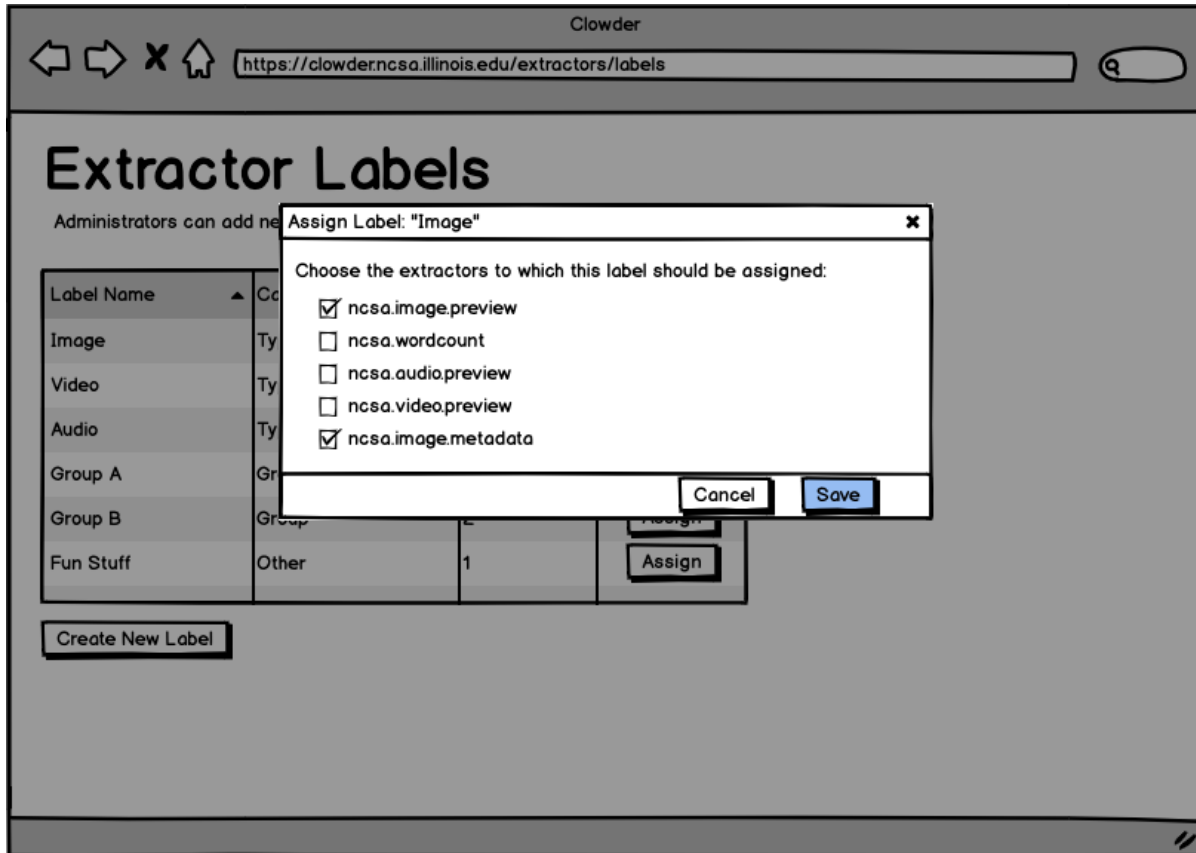
Label Management View



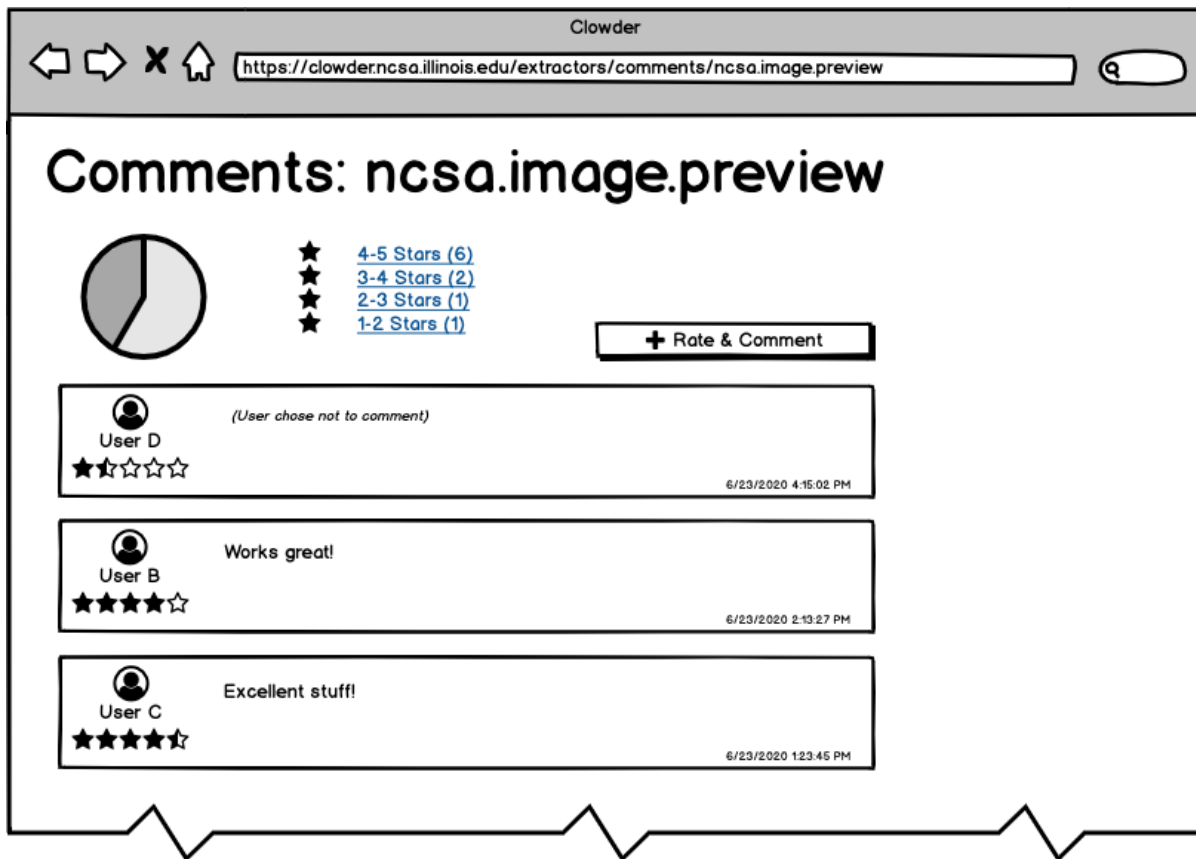
Create New Label



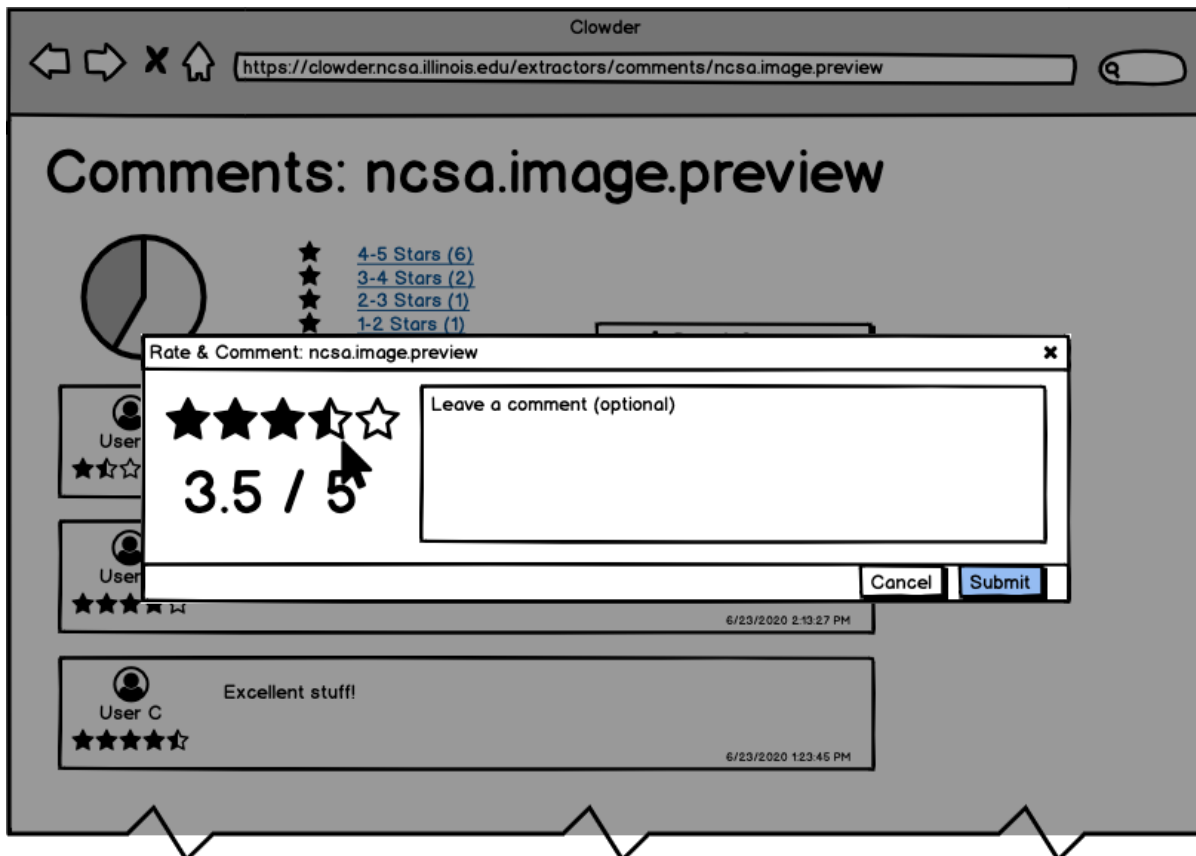
Assign Labels



Comments View



Rate & Comment



History & Metrics View

