extractor_info.json

To run an extractor within a Clowder instance, you will need to create a JSON file called extractor_info.json. This is a flexible file that includes metadata about your extractor that Clowder uses to register the extractor. Once an extractor is registered with Clowder, one can search and submit their extractor on any files or datasets. In addition to simple metadata (e.g., extractor name, extractor author, and extractor description), the extractor_info.json file includes information that tells Clowder how to submit your extractor (i.e., when to submit your extractor and the unit of processing) and information on parameters, if any, that the extractor needs for submission.

Here is a sample extractor_info.json file:

```
{
        "@context": "http://clowder.ncsa.illinois.edu/contexts/extractors.jsonld",
        "name": "ncsa.wordcount",
        "version": "2.0",
        "description": "WordCount extractor. Counts the number of characters, words and lines in the text file
that was uploaded.",
        "author": "Rob Kooper <kooper@illinois.edu>",
        "contributors": [],
        "contexts": [
                {
                "lines": "http://clowder.ncsa.illinois.edu/metadata/ncsa.wordcount#lines",
                "words": "http://clowder.ncsa.illinois.edu/metadata/ncsa.wordcount#words",
                "characters": "http://clowder.ncsa.illinois.edu/metadata/ncsa.wordcount#characters"
                }
        ],
        "repository": [
                {
                "repType": "git",
                "repUrl": "https://opensource.ncsa.illinois.edu/stash/scm/cats/pyclowder.git"
                }
        ],
        "process": {
                  "file": [
                          "text/*",
                          "application/json"
                  ]
        },
        "external_services": [],
        "dependencies": [],
        "labels": [],
        "bibtex": []
}
```

contexts

This field can be left empty, but it is the place where you can define what the metadata fields returned from your extractor mean.

process

Controls when/how an extractor is submitted based on Clowder events. To see a full list of event types refer to Extractor Basics. In the example above, the word count extractor is processed manually on a single file. Below is an example of starting an extractor automatically when a file is added to a dataset:

```
"process": {
        "dataset": [
                "file.added"
                ]
},
```

dependencies

a list of packages the extractor depends upon

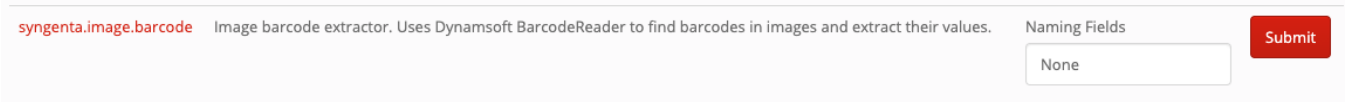
```
"dependencies": [
        "imagemagick",
        "ufraw-batch"
],
```

parameters

Allows parameters to be passed to extractor upon submission.

```
"parameters":{
        "schema": {
                "fields": {
                        "type": "string",
                        "title": "Naming Fields",
                        "default": "None"
                }
        },
        "form": [
                {
                "key": "fields",
                "notitle":false
                }
        ]
},
```

The above code translates as the following within Clowder:



A more complicated example:

```
"parameters": {
        "schema": {
                "orthophoto_resolution": {
                        "type": "string",
                        "title": "Quality",
                        "enum": [
                                "2",
                                "5",
                                "10"
                        ],
                        "default": "5"
                },
        "fast-orthophoto": {
                "type": "boolean",
                "title": "Use sparse reconstruction (fast orthophoto)",
                "default": true
                }
        },
        "form": [
                {
                "key": "orthophoto_resolution",
                "type": "select",
                "notitle": true,
                "titleMap": {
                        "2": "High Quality",
                        "5": "Medium Quality",
                        "10": "Low Quality"
                }
                },
                {
                "key": "fast-orthophoto",
                "inlinetitle": "Use sparse reconstruction (fast orthophoto flag)",
                "notitle": true
                }
        ]
},
```