

# Unified Data Catalog for Imaging Analytics

## Scenario 1: Initialize Catalog with Metadata

Frank Lee

<https://github.com/drfranklee/demo-udc-aircraft>



# Data source management

[Connections](#)[Scan history](#)[Discover databases](#)[Add connection](#)

Connection name	Connection type	Cluster	Data source	Site	State	Scan status	Next scan
UDCt-ClockTower-ScaleS3	S3	http://129.40.101.35	BreaKHis_v1	ClockTower	● Online		
UDCt-haproxy-rhceph01-rgw0	S3	https://9.11.221.111:443	udcuser-bucket		● Online	✓	
UDCt-rhceph01-rgw0	S3	http://rhceph-01.storage-ocp.tuc.stglabs.ibm.com:8080	udc-RGW-bucket		● Online	✓	
UDCt-rhceph03-rgw0	S3	rhceph-03.storage-ocp.tuc.stglabs.ibm.com	udc-RGW3-bucket		● Online		
UDSt-CoS	IBM COS	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	udc-vault	Tucson	● Online	✓	
UDSt-SpectrumScale	Spectrum Scale	scale-bm.storage-ocp.tuc.stglabs.ibm.com	bm_filesystem_2		● Online		

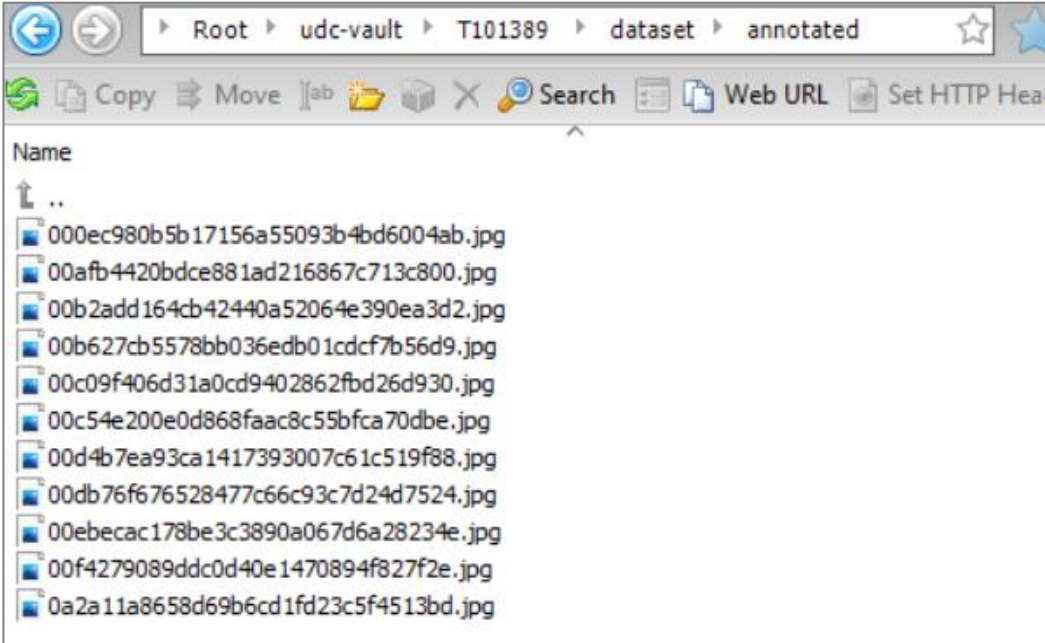
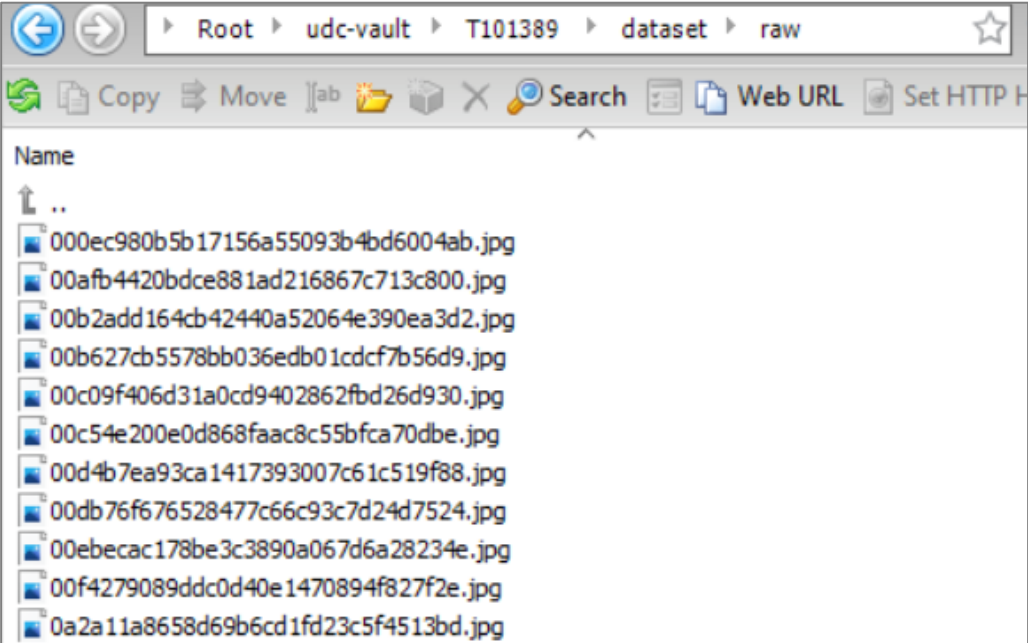
Items per page: 20 ▾

1–6 of 6 items

1 ▾ of 1 pages



Upload Original Data to Cloud Object Storage Bucket



# Query builder

Visual query

SQL query

Select \* from Spectrum Discover where...



filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')

Group by

Choose groupings



Limit results

10000



Setting a lower limit may increase speed of results.

Sort by

e.g. owner asc

Set a sort order to choose which results are returned first.

Reset query

View results

# Query Results

Results view: ?

Condensed

Individual

Inspecting records  
22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

Matched 22 records from metaocean\_view table in 0.084276 seconds

<input type="checkbox"/>	path	filename	udc dem2	u2 type	u2 id	udc dem3	u2 width
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	NULL	NULL	NULL	NULL	NULL
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	NULL	NULL	NULL	NULL	NULL

Items per page: 10

1-10 of 22 items

1 of 3 pages

Create the autotag policy in json

```
{
  "pol_id": "T101389_autotag_u2source_pol",
  "action_id": "AUTOTAG",
  "action_params": {
    "rule": {
      "name": "setFromExistingField",
      "action": "regex",
      "existingField": "filename",
      "regexPattern": "[^/]+",
      "matchNo": 3,
      "newField": "u2-source"
    }
  },
  "schedule": "NOW",
  "pol_filter": "filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')",
  "pol_state": "active",
  "explicit": "true",
  "collections": []
}
```

Execute the policy on server

```
[moadmin@discover1 T101389]$ pwd
/home/moadmin/T101389
[moadmin@discover1 T101389]$ ls
T101389_aircraft_import_scenario1.json  T101389_autotag_u2source.json
T101389_aircraft_import_scenario2.json
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$ gettoken
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload   Total   Spent    Left   Speed
0         0     0      0      0      0      0  0
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$
[moadmin@discover1 T101389]$ curl -k -H "Authorization: Bearer ${TOKEN}" https://localhost/policyengine/v1/policies/T101389_autotag_u2source_pol -X POST -d @./T101389_autotag_u2source.json -H "Content-Type: application/json"
Policy 'T101389_autotag_u2source_pol' added[moadmin@discover1 T101389]$
```

# Tag management

Tags Policies Policy history

Q Add Policy +

Policy	Type	Schedule (UTC)	State	State	Collections	Last modified by	Last modified
ap_cs_body	CONTENTSEARCH	Done	● Active	✔ Complete		sdadmin	9/17/2021, 2:51:16 PM
ap_cs_meta	CONTENTSEARCH	Done	● Active	✔ Complete		sdadmin	9/16/2021, 1:18:06 PM
T101389_autotag_u2source_pol	AUTOTAG	Done	● Active	✔ Complete		sdadmin	9/21/2021, 2:32:38 PM
u3-iphone-gps	CONTENTSEARCH	Done	● Active	✔ Complete		sdadmin	9/17/2021, 12:35:42 AM

Items per page: 20 ▾

1–4 of 4 items

1 ▾ of 1 pages



# Tag management

- Tags
- Policies
- Policy history

Policy name	Type	Started	Ended	Status	Total records	Successful	Skipped	Failed
T101389_autotag_u2source_pol	AUTOTAG	9/21/2021, 2:32:39 PM	9/21/2021, 2:32:41 PM	✔ Complete	22	22	0	0
T101389_contentsearch_u2source_pol	AUTOTAG	9/21/2021, 2:27:44 PM	9/21/2021, 2:27:46 PM	✔ Complete	22	22	0	0
T101389_aircraft_import_scenario2_pol	IMPORT_TAGS	9/21/2021, 1:03:52 AM	9/21/2021, 1:04:11 AM	✔ Complete	22	22	0	0
T101389_aircraft_import_scenario1_pol	IMPORT_TAGS	9/20/2021, 11:21:32 PM	9/20/2021, 11:21:49 PM	✔ Complete	11	11	0	0
T101389_aircraft_import_scenario1_pol	IMPORT_TAGS	9/20/2021, 11:17:06 PM	9/20/2021, 11:17:22 PM	✔ Complete	11	11	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:18:41 PM	9/20/2021, 9:18:59 PM	✔ Complete	11	11	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:16:06 PM	9/20/2021, 9:16:13 PM	✔ Complete	0	0	0	0
T101389_aircraft_import_pol	IMPORT_TAGS	9/20/2021, 9:08:05 PM	9/20/2021, 9:08:40 PM	✔ Complete	0	0	0	0
an cs body	CONTENTSEARCH	9/17/2021,	9/17/2021,	✔ Complete	16	16	0	0



Metadata Tag “u2-source” Populated with Values Extracted by AutoTag Policy Run

Results view: ⓘ

Matched 22 records from metaocean, view table in 0.082699 seconds

Condensed

Individual

Inspecting records  
22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

<input type="checkbox"/>	path	filename	u2 source
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdf7b56d9.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00ebecac178be3c3890a067d6a28234e.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00f4279089ddc0d40e1470894f827f2e.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/0e2a11a8658d69b6cd1fd23c5f4513bd.jpg	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/raw/000ec980b5b17156a55093b4bd6004ab.jpg	raw
<input type="checkbox"/>	udc-vault/	T101389/dataset/raw/00afb4420bdce881ad216867c713c800.jpg	raw
<input type="checkbox"/>	udc-vault/	T101389/dataset/raw/00b2add164cb42440a52064e390ea3d2.jpg	raw
<input type="checkbox"/>	udc-vault/	T101389/dataset/raw/00b627cb5578bb036edb01cdf7b56d9.jpg	raw
<input type="checkbox"/>	udc-vault/	T101389/dataset/raw/00c09f406d31a0cd9402862fbd26d930.jpg	raw

Items per page: 50

1-22 of 22 items

1 of 1 pages

Query builder

filename LIKE ('%T101389%') AND filetype LIKE ('%jpg%')



Results view: ?

Condensed Individual

Inspecting records  
22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

Matched 22 records from *indoocean\_view* table in 0.082699 seconds

22 items selected				Generate report	Add tags	Export data	Cancel
<input checked="" type="checkbox"/>	path	filename	u2 source				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	annotated				
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	annotated				
Items per page: 50 1-22 of 22 items				1	of 1 pages		

Results view: ?

Condensed

Individual

Inspecting records

22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

# Add tags

22 files selected

udc-dem2



T101389-s1



[Add another tag](#)

Cancel

Submit

<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg

Items per page: 50 1-22 of 22 items

1 of 1 pages

Results view: ?

Condensed

Individual

Inspecting records

22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

Matched 22 records from metaocean\_view table in 0.05813 seconds

<input type="checkbox"/>	path	filename	udc dem2	u2 source
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	T101389-s1	annotated
Items per page: 10			1 of 3 pages	
1-10 of 22 items				

# Query builder

Visual query   SQL query

Query preview...

Temperature 0/1

Tier 0/2

Timesinceaccess 0/5

U2-id 0/1

U2-source 0/3

U2-type 0/1

Udc-dem2 0/4

Udc-dem3 0/1



<input type="checkbox"/> udc-dem2	count
<input type="checkbox"/> NULL	21,958
<input type="checkbox"/> T101389-s1	22
<input type="checkbox"/> E1016270-SGDSO	8
<input type="checkbox"/> E1016270-demo	1

Items per page: 10

1-4 of 4 items

1 of 1 pages

# Query Results

Results view:

Condensed

Individual

Displaying records ⓘ

22

Condensed results

1

Grouped By

☒ Udc-dem2

Show more...

Grouped 22 results from mrcapacity table in 0.268984 seconds

☐ udc dem2

☐ T101389-s1

Total Files		Total Size	
22		5.33 MiB	
Items per page:	10 ▾	1-1 of 1 items	1 ▾ of 1 pages

Query builder

udc-dem2 IN ( 'T101389-s1' )



Query Results

Results view:

Condensed

Individual

Inspecting records  
22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

Matched 22 records from metaocean\_view table in 0.0743 seconds

22 items selected					Generate report	Add tags	Export data	Cancel
✓	path	filename	filetype	udc dem2	u2 source			
✓	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdcf7b56d9.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00ebecac178be3c3890a067d6a28234e.jpg	jpg	T101389-s1	annotated			
✓	udc-vault/	T101389/dataset/annotated/00f4279089ddc0d40e1470894f827f2e.jpg	jpg	T101389-s1	annotated			
Items per page: 10					1-10 of 22 items		1	of 3 pages

Query Results

Results view:

Condensed

Individual

Inspecting records

22

Grouped By

☐ Cluster

☐ Datasource

☐ Filegroup

☐ Filename

☐ Fileset

Show more...

22 items selected

Generate reportAdd tagsExport dataCancel

<input checked="" type="checkbox"/>	path	filename	udc dem2	u2 source
			T101389-s1	annotated
			T101389-s1	annotated
			T101389-s1	annotated
			T101389-s1	annotated
			T101389-s1	annotated
			T101389-s1	annotated
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	T101389-s1	annotated
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	T101389-s1	annotated
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	T101389-s1	annotated
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00ebecac178be3c3890a067d6a28234e.jpg	T101389-s1	annotated
<input checked="" type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00f4279089ddc0d40e1470894f827f2e.jpg	T101389-s1	annotated

Items per page: 101–10 of 22 items1 of 3 pages

Name

T101389-s1-manifest-v20210921

Query Term: udc-dem2 IN ('T101389-s1')

Cancel

Submit



# Reports

✔

Report has been Created

The report T101389-s1-manifest-v20210921 was created

✕

Report	Last run	Duration	Status	Output size
T101389-s1-manifest-v20210921	Tue Sep 21 2021 14:49:17 GMT-0500 (Central Daylight Time)	0 seconds	Complete	10.32 KiB
demo-csv	Thu Sep 09 2021 18:39:49 GMT-0500 (Central Daylight Time)	0 seconds	Complete	2.34 KiB
E101475_d1-list	Thu Apr 29 2021 08:16:33 GMT-0500 (Central Daylight Time)	0 seconds	Complete	8.17 KiB
E1014753-data-manifest	Thu Apr 29 2021 00:36:52 GMT-0500 (Central Daylight Time)	0 seconds	Complete	8.08 KiB
BH01 Dataset Report for FS21EMEA	Tue Feb 09 2021 00:14:30 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.75 MiB
TD-BH01 Report	Sat Jan 30 2021 18:15:17 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.73 MiB
BreakHist Summary	Sat Jan 30 2021 17:56:08 GMT-0600 (Central Standard Time)	0 seconds	Complete	3.33 MiB
Age Report Detail. 90-180 Days	Tue Dec 22 2020 02:24:47 GMT-0600 (Central Standard Time)	0 seconds	Complete	308 Bytes
Age Report Summary. 90-180 Days	Tue Dec 22 2020 02:24:47 GMT-0600 (Central Standard Time)	0 seconds	Complete	41 Bytes
Age Report Detail. 720+ Days	Tue Dec 22 2020 02:24:46 GMT-0600 (Central Standard Time)	0 seconds	Complete	308 Bytes

## Manifest Can be Downloaded Locally as CSV File

[illegible]

Condensed

Individual

Inspecting records  
22

Grouped By

- ☐ Cluster
- ☐ Datasource
- ☐ Filegroup
- ☐ Filename
- ☐ Fileset

Show more...

<input type="checkbox"/>	path	filename	cluster	udc dem2	u2 source
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/000ec980b5b17156a55093b4bd6004ab.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00afb4420bdce881ad216867c713c800.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b2add164cb42440a52064e390ea3d2.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00b627cb5578bb036edb01cdf7b56d9.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c09f406d31a0cd9402862fbd26d930.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00c54e200e0d868faac8c55bfca70dbe.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00d4b7ea93ca1417393007c61c519f88.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00db76f676528477c66c93c7d24d7524.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00ebecac178be3c3890a067d6a28234e.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
<input type="checkbox"/>	udc-vault/	T101389/dataset/annotated/00f4279089ddc0d40e1470894f827f2e.jpg	df3b7ae7-ab82-756b-01c4-c9f3ac8fae42	T101389-s1	annotated
Items per page: 10 1-10 of 22 items				1 of 3 pages	

Query builder

udc-dem2 like ('%T101389-s1%')



