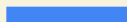


datalad-registry



<http://registry.datalad.org>

<https://github.com/datalad/datalad-registry>

Goal

1. To allow discovery of data from a large collection of datasets, Datalad datasets and git repos touched by ``datalad run``
2. To allow discovery of reuses of datasets as a submodule of a super dataset

Automation as the Defining Principle

- In discovery of datasets
- In extraction of metadata from the datasets
- Updating of the dataset state and metadata

Current User View

Providing up-to-date info on 12,000+ datasets

1. All datasets on <https://datasets.datalad.org/>
2. Datasets in the wild as they are discovered by <https://github.com/datalad/datalad-usage-dashboard>
 - a. Datasets that reside on GitHub
 - b. Datasets that reside on GIN
3. Information of each dataset includes metadata of the dataset
 - a. "metalad_core"
 - b. "metalad_studyminimeta"
 - c. "datacite_gin"
 - d. "bids_dataset"
 - e. "dandi" (not from datalad-metalad - just loaded JSON dumps)

Current User View

URL	Dataset	Commit	Annex keys	Annexed files in working tree	Size of Annexed files in working tree (bytes)	Last update	.git/objects Size (KiB)	Metadata
https://gin.g-node.org/msz/2023_Stokowski_Perioinfarex	4ac2694c-bfff-48ce-8052-cdb6285fc2e7	f06b296e96	0	12777	67708876219	2024-01-28T01:09:23+0000	4281391	metalad_core , datacite_gin
https://github.com/bats-core/bats-core.git	v1.11.0-RC1	v1.11.0-RC1				2024-01-28T00:30:27+0000	3145	
https://github.com/janelia-flyem/dvid.git	v1.0.0-33-g84b72ab6	v1.0.0-33-g84b72ab6				2024-01-28T00:28:27+0000	16580	
https://github.com/conda-forge/conda-forge-repodata-patches-feedstock.git	9f57397	9f57397				2024-01-28T00:14:27+0000	1698	
https://github.com/paramiko/paramiko.git	3.4.0-9-geb470b12	3.4.0-9-geb470b12				2024-01-28T00:07:27+0000	6942	
https://github.com/tomerfiliba/org-rpyc.git	5.3.1-35-g19a9f03	5.3.1-35-g19a9f03				2024-01-27T23:57:26+0000	3488	
https://github.com/PySlurm/pyslurm.git	v23.11.0	v23.11.0				2024-01-27T23:26:25+0000	5944	
https://github.com/bacalhau-project/bacalhau.git	v1.2.1-rc2-94-gdc353990	v1.2.1-rc2-94-gdc353990				2024-01-27T23:20:25+0000	185865	
https://github.com/saulpau/visidata.git	v3.0-82-g5ec117d9	v3.0-82-g5ec117d9				2024-01-27T23:04:24+0000	44496	
https://gin.g-node.org/felixh/CoRR_cat12.8_1mm	87442981-968f-4ed7-958f-b4d620b6babb	6cf67ed4933	3	236064	1633805624330	2024-01-27T18:03:23+0000	252886	metalad_core
https://gin.g-node.org/AIDAqc/datasets/7_h_RT_He	810318a3-284c-4f23-a73d-b0c7015e4ff2	8bf0826	30	30	1211525408	2024-01-27T17:34:51+0000	83657	metalad_core
https://gin.g-node.org/lucasgautheron/trading_zones_material	3a7227ea-d9a3-41bc-ae19-206924e5a111	0735265	120	50	6436731421	2024-01-27T17:32:44+0000	57709	datacite_gin , metalad_core
https://gin.g-node.org/doi/2023_Blaschke_Stroke_IDCS_rsfMRI	83b73949-b6c3-4b2b-81b2-c3f7236a0848	10.12751/g-node.nbmq9r	33239	40158	93516897516	2024-01-27T17:32:39+0000	30263	datacite_gin , metalad_core
https://gin.g-node.org/Aswendt_Lab/CRC_WPI_raw_data	5fb584e9-5e1d-4b73-938b-9ff53a384ee7	5ffad8c5aa	47008	63191	356769317935	2024-01-27T17:31:58+0000	49221	metalad_core
https://gin.g-node.org/koch_means_cook/pedr1-derivatives	ec0a0df6-811f-4c80-aa7e-b0fe8ed4c999	4413be5e6c	28218	3673	4751892830	2024-01-27T17:29:30+0000	38560	metalad_core , datacite_gin
https://gin.g-node.org/gallantlab/story_listening	92bd3b0a-8102-45fd-8ff3-b4c5a13ed7c3	e80c315	38	38	30402661787	2024-01-27T17:29:21+0000	36961	metalad_core , datacite_gin
https://gin.g-node.org/labgas/proj_erythritol_4a-secondlevel	d5412b56-8d98-40bf-95ad-a09f74961ff	079e4c4ff	8873	5773	162726151668	2024-01-27T17:28:50+0000	28453	metalad_core
https://gin.g-node.org/felixh/d003949_ReproVBM	16088113-c4da-437e-a216-7a38acf638ef	53eac3d08	13078	21016	43787036458	2024-01-27T17:27:43+0000	21417	metalad_core
https://gin.g-node.org/felixh/uwmadison	e3975480-9982-48a8-9ea1-403e4a913de5	682e5acdd	8250	12096	167361584206	2024-01-27T17:26:57+0000	13642	metalad_core
https://gin.g-node.org/arefks/AIDAqc_test_data	f1a7ead6-a448-4c29-aad5-921e59db6aba	9461caf64	8626	12080	36887894073	2024-01-27T17:26:08+0000	10741	datacite_gin , metalad_core

1 - 20 of 12715

Page: 1 2 3 4 5 ... 635 636

Current User View

A simple interface that provides basic search

haxby Filter URLs [remove filter](#)

URL	Dataset	Commit	Annex keys	Annexed files in working tree	Size of Annexed files in working tree (bytes)	Last update	.git/objects Size (KiB)	Metadata
https://github.com/OpenNeuroDatasets-JSONLD/ds001297.git	2e429bfe-8862-11e8-98ed-0242ac120818	00001-4-g2c4c2f7	0	495	10593648758	2024-01-16T17:47:49+0000	1045	metadad_core , bids_dataset
https://github.com/OpenNeuroDatasets-JSONLD/ds000233.git	aa38f5a2-89f6-11e8-abf3-0242ac120804	1.0.1-4-g3a10cc7	0	362	5090155130	2024-01-16T17:41:23+0000	922	metadad_core , bids_dataset
https://datasets.datalad.org/centerforopenneuroscience/nuisance/data/dbic/bids/.git	b3e8cdae-3608-11e9-901b-002590f97d84	cac6fee7	0	2960	10169645676	2023-11-24T06:40:15+0000	2130	metadad_core , bids_dataset
https://datasets.datalad.org/openneuro/ds001297.git	2e429bfe-8862-11e8-98ed-0242ac120818	00001	0	495	10593648758	2023-11-24T05:40:14+0000	1065	metadad_core , bids_dataset
https://datasets.datalad.org/openneuro/ds000233.git	aa38f5a2-89f6-11e8-abf3-0242ac120804	1.0.1	0	362	5090155130	2023-11-24T05:38:53+0000	882	bids_dataset , metadad_core
https://datasets.datalad.org/openneuro/ds000241.git	78c36e78-8927-11e8-a8c0-0242ac120e26	00002	0	270	4971868833	2023-11-24T05:38:49+0000	620	bids_dataset , metadad_core
https://datasets.datalad.org/openneuro/ds000105.git	c01c2348-8708-11e8-938d-0242ac120804	3.0.0	0	93	1879627119	2023-11-24T05:38:13+0000	226	bids_dataset , metadad_core
https://datasets.datalad.org/openfmr/ds000241/.git	f7fc7f8c-c436-11e7-af7c-002590f97d84	1.0.0+1-3-gd81d8fc	0	535	4972575726	2023-11-24T05:35:59+0000	813	metadad_core , bids_dataset
https://datasets.datalad.org/openfmr/ds000237/.git	f733d2f6-49f1-11e7-bf68-003048ce7f98	1.0.0+1-3-gb5b7b97	146	801	5007376378	2023-11-24T05:35:55+0000	1627	bids_dataset , metadad_core
https://datasets.datalad.org/openfmr/ds000105/.git	0db8aed8-7df7-11e6-8819-002590f97d84	2.0.2+3-3-g4928cd5e2	4	166	1874749325	2023-11-24T05:35:53+0000	19443	metadad_core , bids_dataset
https://datasets.datalad.org/labs/haxby/life/derivatives/fmriprep-1.5.0/.git	f47cde46-df31-11e9-9c4f-002590f97d84	c72653077	8051	8996	161885598098	2023-11-24T05:35:09+0000	12479	metadad_core
https://datasets.datalad.org/labs/gobbini/famface/data/.git	699b6994-97f9-11e7-bf68-003048ce7f98	81D5-v1.0.1-3-ge59ba76b	487	546	10969854355	2023-11-24T05:34:46+0000	4981	metadad_core , bids_dataset
https://datasets.datalad.org/labs/haxby/attention/derivatives/afni/.git	b5578448-15a5-11e8-8fe5-003048ce7f98	04e700f1	3744	6792	92251518004	2023-11-24T05:34:44+0000	5689	metadad_core
https://datasets.datalad.org/labs/haxby/life/containers/.git	bd2e63c2-62c1-11e9-82b0-5254a004089c	0.2-170-g9c9d275f	2	43	92298287861	2023-11-24T05:34:40+0000	2215	metadad_core
https://datasets.datalad.org/labs/haxby/riders/.git	6eeadb86-84be-11e6-b24c-002590f97d84	bids-2.0.0	557	178	28168593269	2023-11-24T05:34:36+0000	1828	bids_dataset , metadad_core
https://datasets.datalad.org/labs/haxby/attention/.git	73f3d2f6-49f1-11e7-bf68-003048ce7f98	v1.1.0-21-ga00a6d0	208	232	4456501077	2023-11-24T05:34:36+0000	780	metadad_core , bids_dataset
https://datasets.datalad.org/labs/haxby/attention/derivatives/fmriprep/.git	c18c0066-15a5-11e8-8fe5-003048ce7f98	d816eed	532	544	9443275887	2023-11-24T05:34:32+0000	976	metadad_core
https://datasets.datalad.org/labs/haxby/life/.git	4119b6c6-c32a-11e7-8fe5-003048ce7f98	nondefaced-28-g47d39b0	115	115	6462016470	2023-11-24T05:34:32+0000	1105	bids_dataset , metadad_core
https://datasets.datalad.org/labs/haxby/attention/derivatives/freesurfer/fsaverage6/.git	8d72cae0-15a6-11e8-8fe5-003048ce7f98	bb422bd	178	187	604539178	2023-11-24T05:34:32+0000	311	metadad_core
https://datasets.datalad.org/labs/haxby/attention/derivatives/pymvpa/.git	88d5a2f0-15a5-11e8-8fe5-003048ce7f98	c112322	104	104	1808696737	2023-11-24T05:34:31+0000	202	metadad_core

Current User View

The search interface can also be used to search for forks and reuses of a dataset.

b02e63c2-62c1-11e9-82b0-52540040489c						
URL	Dataset	Commit	Annex keys	Annexed files in working tree	Size of Annexed files in working tree (bytes)	Last update
https://datasets.datalad.org/repromim/containers/.git	b02e63c2-62c1-11e9-82b0-52540040489c	0.20230525.0-184-g05b6052a	248	398	1405747452669	2024-01-23T18:45:20+0000 2509
https://github.com/ReproNim/containers.git	b02e63c2-62c1-11e9-82b0-52540040489c	0.20230525.0-184-g05b6052a	0	398	1405747452669	2024-01-23T11:28:02+0000 3776
https://gin.g-node.org/adswa/bidsdata	57165da8-cf64-4aa9-b8eb-bf79d8dccb10	7e5abbf	2	2	49861132	2024-01-16T17:52:36+0000 408
https://gin.g-node.org/melanieganz/BrainHackNordic2022	134b0c87-e325-4f8a-be0e-88a31f632025	1f72ebb9	2	2	49861190	2024-01-16T17:49:55+0000 416
https://gin.g-node.org/adswa/bids-data	68bdb3f3-eafa-4a48-bddd-31e9d4e8b8242	345c975	2	2	49861839	2024-01-16T17:48:36+0000 428
https://gin.g-node.org/adswa/processing-workflow-tutorial	47a9ef9e-9fa0-49cd-873f-c51cc501da96	efbb4a6	424	504	7103481914	2024-01-16T17:43:44+0000 723
https://gin.g-node.org/jsheunis/bids-data	894d66ee-441a-4101-9620-fa758b7461c7	b277d03	2	2	49863184	2024-01-16T17:41:52+0000 416
https://github.com/courtois-neuromod/hcpnrt.fmrprep.git	167b1afc-ab5f-473b-932b-7e7a28c37dc3	0.5.0-5-g206ad779	0	11868	451983983937	2024-01-09T14:10:55+0000 25819
https://github.com/courtois-neuromod/movie10.fmrprep.git	3f0351b8-6d9f-4d05-841e-948che830bc7	2693619b0	0	6612	627031710539	2024-01-09T14:07:45+0000 26710
https://github.com/courtois-neuromod/shinobi.fmrprep.git	99a5a823-59ed-4f4e-9fbd-971bb5e7d1d5	0.10.0-7-gf6670bb88	0	3904	424300057373	2024-01-09T14:01:31+0000 19969
https://github.com/courtois-neuromod/harrypotter.fmrprep.git	e7c71cca-4d8e-4bd1-b8c7-5fe43e44c1b9	2321aeba	0	620	71116533441	2024-01-09T13:59:21+0000 51965
https://datasets.datalad.org/openneuro-derivatives/ds001734-fmrprep/.git	bc91dbbd-44bf-4840-9865-b7d666c2b8dc	d0cbf2f7fb	0	52396	1993403874396	2023-11-24T06:45:26+0000 77579
https://datasets.datalad.org/openneuro-derivatives/ds000243-fmrprep/.git	4d22bf624-9003-4289-95d7-db14778c334	544965de61	0	48316	222127733932	2023-11-24T06:44:56+0000 72967
https://datasets.datalad.org/openneuro-derivatives/ds003338-fmrprep/.git	50889575-4cb6-4608-9f27-1c92dbc6abbd	e021193e7	0	12973	669841192522	2023-11-24T06:43:58+0000 13553
https://datasets.datalad.org/openneuro-derivatives/ds002898-fmrprep/.git	dea9fde4-42c6-488a-a06d-333c5a17cd31	fb9bf287a	0	15145	200585778494	2023-11-24T06:43:54+0000 24448
https://datasets.datalad.org/openneuro-derivatives/ds003097-fmrprep/.git	adb04e29-bf86-4d45-80d2-50a253d058a6	5c8121e83	0	4587	70456456213	2023-11-24T06:43:51+0000 16185
https://datasets.datalad.org/openneuro-derivatives/ds000208-fmrprep/.git	c0eed5ef-a2dd-4b6e-84e7-d9b9f6ae37f8	88375b7958	0	29123	120746955906	2023-11-24T06:43:47+0000 51918
https://datasets.datalad.org/openneuro-derivatives/ds002338-fmrprep/.git	271805b6-e034-4731-8817-6c266b111643	1b6f75639	0	9166	90664058310	2023-11-24T06:43:43+0000 16621
https://datasets.datalad.org/openneuro-derivatives/ds000212-fmrprep/.git	8e0fb7b-2e1f-4b12-ad8a-bababb87fa72	e64b04c82c	0	25182	263351987544	2023-11-24T06:43:37+0000 42771
https://datasets.datalad.org/openneuro-derivatives/ds002322-fmrprep/.git	8fa166e5-20d4-4cb6-8f59-155a5112d508	9541233eb3	0	10315	47874817625	2023-11-24T06:43:33+0000 38085

Current Developer View

An API with OpenAPI V3 documentation at <http://registry.data-lad.org/openapi/>

- The OpenAPI V3 documentation is suitable for human consumption through Swagger, ReDoc, or ReapiDoc
- The OpenAPI V3 documentation can be used by other parties generate code to access the API
- Supported operations
 - Registering new datasets (only on typhon, not on public registry.data-lad.org)
 - Querying for datasets based on restrictions
 - Fetching datasets and metadata by internal IDs

Current Developer View

☐	Name	Image	Status	Port(s)	Last started
☐	 datalad-registry		Running (7/7)		3 minutes ago
☐	 db-1 099cd16d0017	docker.io/postgres:late	Running	5432:5432 	3 minutes ago
☐	 broker-1 1b8e905c4942	docker.io/rabbitmq:3-m	Running	15672:15672  Show all ports (2)	3 minutes ago
☐	 backend-1 48b93ed3452c	docker.io/redis:7	Running	6379:6379 	3 minutes ago
☐	 worker-1 1a76693f88f5	datalad-registry:dev	Running		3 minutes ago
☐	 web-1 b230078609c7	datalad-registry:dev	Running	5000:5000 	3 minutes ago
☐	 monitor-1 ace254f658d1	datalad-registry:dev	Running	5555:5555 	3 minutes ago
☐	 scheduler-1 8a9c250d3e4d	datalad-registry:dev	Running		3 minutes ago





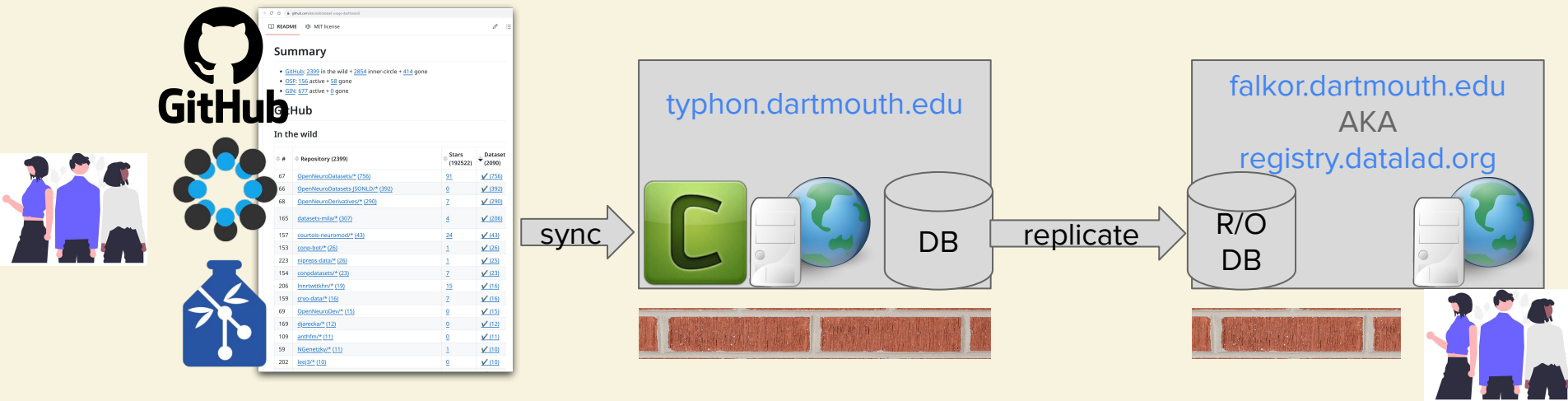



Celery Related Services

Current Developer View

Ability to launch a **read-only instance** that consists of only the web and DB services

- Protect the main instance depicted in the previous slide from abuse
- Multiple read-only instances can split the workload
- A read-only instance is set up to be in sync with the main instance



Ongoing Development

Styling facelift by Stephan: [datalad-registry/issues/290](https://github.com/datalad/datalad-registry/issues/290)

< > G File /Users/jsheunis/Documents/psynif/dlreg/index.html

[☆](#) [▼](#) [🔄](#) | [🔍](#) [📄](#) [👤](#) [Relaunch to update](#)



registry

Dataset URL	Dataset ID	Commit	Nr of Annex keys	Nr of Annexed files	Size of Annexed files	Last update	Size of .git/objects	Metadata
https://gin.g-node.org/felixh/ds003145_ReproVBM	b06ae36d-167a-4e36-bde7-c371034288b4	42d6e53	0	0	0	2024-01-20T17:24:55+0000	412	metatlad_core
https://gin.g-node.org/felixh/ds003095_ReproVBM	b1c25623-4d29-43f7-87da-51026385e7ba	2a7d16b5	1762	2812	6830687261	2024-01-20T17:24:54+0000	3108	metatlad_core
https://gin.g-node.org/felixh/ds003190_ReproVBM	f360ed2e-140f-41f4-ad91-1c7ca14db5ab	99901b3	0	0	0	2024-01-20T17:24:54+0000	360	metatlad_core
https://gin.g-node.org/felixh/ds003520_ReproVBM	ac f86061-9ade-45a4-828e-a2a773bd99da	49ccd90	0	0	0	2024-01-20T17:24:54+0000	356	metatlad_core
https://gin.g-node.org/felixh/ds003500_ReproVBM	8f86a91b-cc49-4ea8-bdf6-e88661e1dac9	db95dc0	1716	2738	5817735811	2024-01-20T17:24:54+0000	3273	metatlad_core
https://gin.g-node.org/ivis-mizuguchi/datalad-siblings	d4f1f904-1338-42b9-8fb3-8f6eaced9575	3bc4fc0	0	0	0	2024-01-20T17:24:53+0000	92	metatlad_core
https://gin.g-node.org/felixh/ds003552_ReproVBM	cb76dd8d-ae62-4f93-a0fd-8f05c56f5f5d	4ed8676	0	0	0	2024-01-20T17:24:53+0000	360	metatlad_core
https://gin.g-node.org/ivis-mizuguchi/0530-4	822a921b-9369-4d35-9d9d-4091b1dfca7d	41bc966	4	10	12888019	2024-01-20T17:24:53+0000	376	metatlad_core
https://gin.g-node.org/felixh/ds003639_ReproVBM	3e411ba4-ec01-42a5-b14d-e4a4853a0708	b023a9a6	1164	1850	3799639732	2024-01-20T17:24:52+0000	2278	metatlad_core
https://gin.g-node.org/felixh/ds000009_ReproVBM	94732214-8be9-446c-af6b-abdf55fd2bd0	f81d67a9	1118	1776	3871561023	2024-01-20T17:24:52+0000	2036	metatlad_core
https://gin.g-node.org/LAAC-LSCP/longform-hardware-audio-test	4c56d872-1c72-4ce7-89ea-2ddfefe9cf7f	3266037	103	62	224306157	2024-01-20T17:24:51+0000	187	metatlad_core
https://gin.g-node.org/felixh/ds003445_ReproVBM	a1b4b20e-e5e2-4210-b4ec-82a8c1bfc6d4	c074a7c	60	74	147113814	2024-01-20T17:24:51+0000	144	metatlad_core
https://gin.g-node.org/felixh/ds003633_ReproVBM	9167ee01-e7cd-4bd6-988c-344bec17f6c3	d581ac6	520	814	1736361230	2024-01-20T17:24:51+0000	937	metatlad_core
https://gin.g-node.org/felixh/ds003399_ReproVBM	6a6c9688-5da6-46c5-ab58-f6447e086d85	81dbecc	0	0	0	2024-01-20T17:24:50+0000	388	metatlad_core
https://gin.g-node.org/felixh/ds000052_ReproVBM	a84b5ff5-9139-4cc4-8c5f-8874dac045aa	5e1f444	612	962	1788915105	2024-01-20T17:24:50+0000	1254	metatlad_core
https://gin.g-node.org/felixh/ds003439_ReproVBM	2eed8747-991f-4420-b65e-3a41336941b6	2253843	60	74	147113771	2024-01-20T17:24:50+0000	144	metatlad_core
https://gin.g-node.org/felixh/ds003008_ReproVBM	27b86b54-0c0a-4b4c-8352-f72a0c08971a	7eabb7c7	2268	3626	7709699424	2024-01-20T17:24:50+0000	3988	metatlad_core
https://gin.g-node.org/lucasgautheron/simulation-scenarios-rte	3aa1f968-7950-42a8-8f9c-e9321593953b	7e4e3cc	22	6	406606938	2024-01-20T17:24:49+0000	6833	metatlad_core
https://gin.g-node.org/felixh/ds003171_ReproVBM	dad2ba76-d2dc-47a9-8ce9-dbab9c089f6c	ccb581E	796	1258	2736324136	2024-01-20T17:24:46+0000	1403	metatlad_core
https://gin.g-node.org/felixh/ds003195_ReproVBM	f91e5d95-9d0b-4e9a-b09b-38988393c94c	46e08bd	0	0	0	2024-01-20T17:24:45+0000	356	metatlad_core

81 - 100 of 12586
 Page: 1 2 3 4 5 6 7 8 9 ... 629 630

Ongoing Development

Integration with datalad-catalog: [datalad-registry/pull/278](https://github.com/datalad/catalog/pull/278)

- State: one metadata field at a time
- Goal: Provide datalad-catalog landing pages for datasets

Upcoming Development

- Elaborate search scheme (main focus)
- Dataset level metadata from other extractors
- File level metadata extractions
- Anticipated challenge: scalability
 - Picking the right database technology
 - Search through a large amount of semi-structured data and producing result in a reasonable amount of time.