

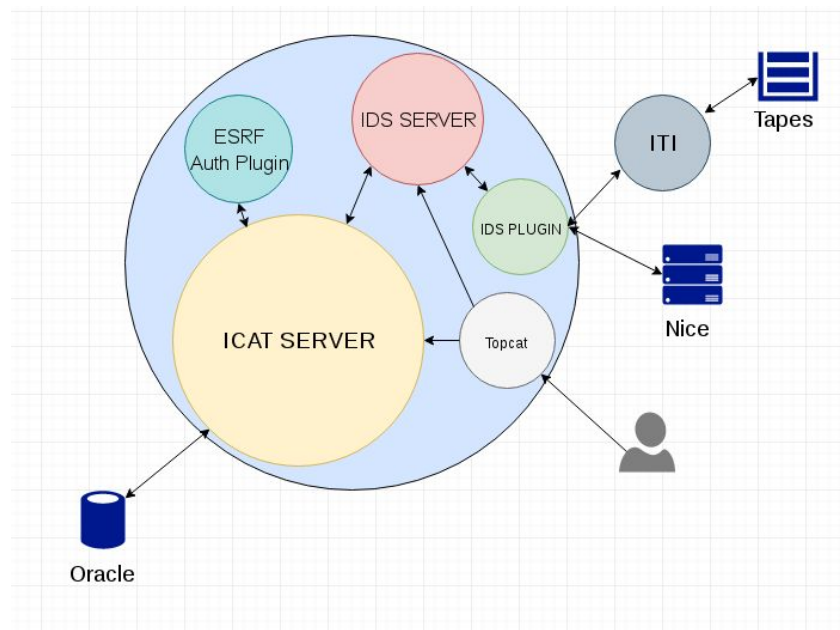
Datahub Demo

Alex de Maria Antolinos
Software Engineer
Data Manager@Data Analysis Unit
Software Group
ESRF
12.11.2017

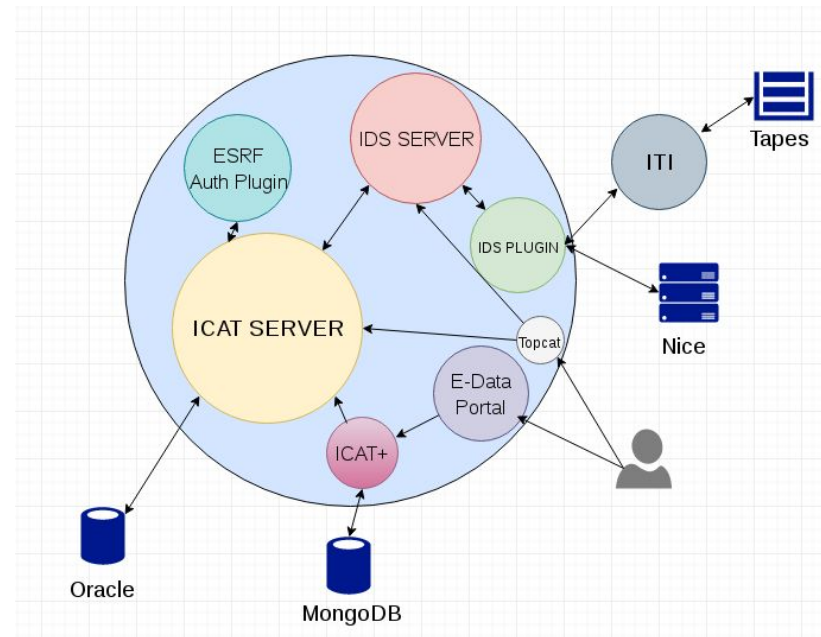
Main activity

- Architecture
- ICAT Plus
 - Overcome ICAT's limitations
 - Live files
 - Complex data structures
 - E-logbook
 - Implementation
 - Integration with ICAT
 - Secure
 - Restricted access to ICAT from outside
 - Improve data retrieval
- Datahub. Frontend done in React
 - Similar to TopCAT
 - + DOI plugin integrated
 - + Electronic logbook
 - +Many "limitations"
 - Unit storage is Dataset
 - Raw data is read-only
 - Statistics are calculated in real time
 - Etc...

Architecture



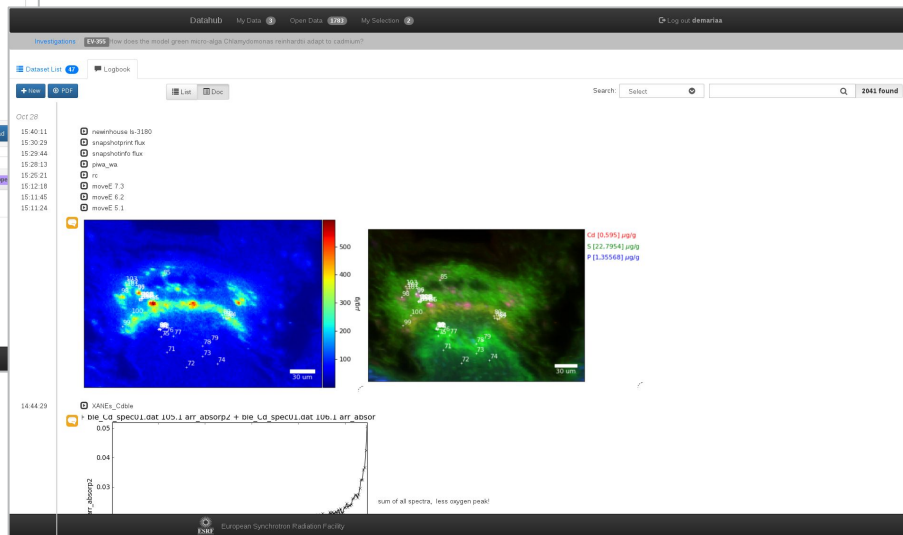
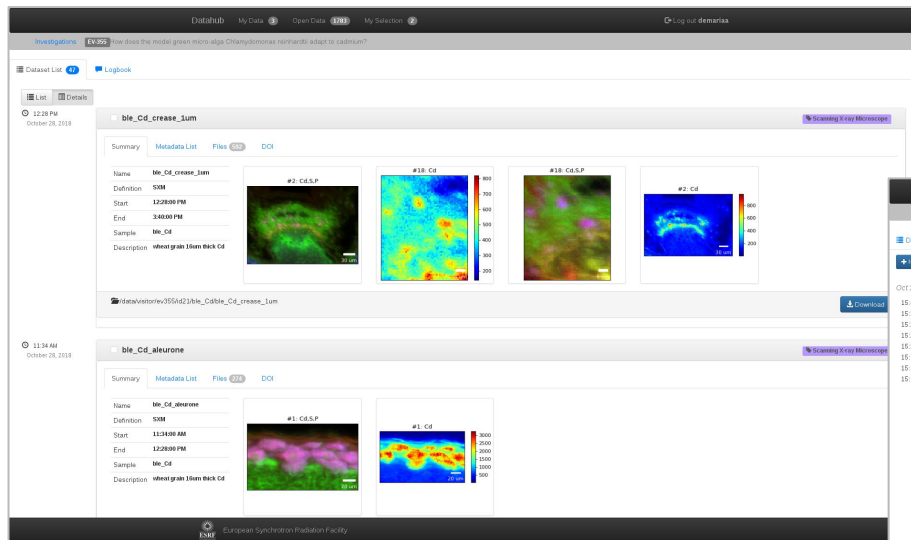
CURRENT



NEW

Overcome some ICAT's "limitations"

- Live files stored on GridFS (mongoDB) and accessible via restful WS
- Storing complex data structures like documents (Example logbook)



API

- Restful web services written in NodeJS
- Swagger (23 entry points)
- No access to entityManager

swagger

ICAT+ 1.0.0

[Base URL: <https://indesaria.esrf.fr:8000/>]

ICAT+ API with swagger <http://swagger.io> or on c.fra.node.net/#swagger.

[Terms of service](#)
[Contact the developer](#)
 MIT
[Find out more about Swagger](#)

Schemes

datacollection ICAT Datacollections Not available for the moment: <http://swagger.io>

GET `/datasets/{datasetId}/datacollection` Gets the datacollections related to a dataset.

events Events form a subset from all the resources managed by ICAT+. Events include annotations (ie messages written by users) and notifications (messages written by programs). Not available for the moment: <http://swagger.io>

POST `/investigations/{investigationName}/instrument/{instrumentName}/events/notification` Creates a new notification

GET `/investigations/{investigationId}/events/pdf` Exports a PDF file with all events

GET `/investigations/{investigationId}/events` Get events associated to an given investigation

POST `/investigations/{investigationId}/events` Creates a new event

POST `/investigations/{investigationId}/events/file` Uploads an Event that contains a file

GET `/investigations/{investigationId}/events/{eventId}/file` Downloads the file of a event

investigation Resources attached to a investigation Not available for the moment: <http://swagger.io>

POST `/investigations/{investigationName}/normalize` Given an investigation name it returns its name normalized

POST `/investigations/{investigationId}/events/query/count` It runs a mongoose query based in find, sort, skip and limit parameters and returns the number of entities retrieved

EntityManager

[https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT dataset FROM Dataset dataset JOIN dataset.parameters parameters JOIN parameters.type parameterType JOIN dataset.investigation investigation where dataset.id = '150396772' include dataset.parameters.type](https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT%20dataset%20FROM%20Dataset%20dataset%20JOIN%20dataset.parameters%20parameters%20JOIN%20parameters.type%20parameterType%20JOIN%20dataset.investigation%20investigation%20where%20dataset.id%20=%20'150396772'%20include%20dataset.parameters.type)

Data size: 64K

```
{
  "Dataset": {
    "id": 150396772,
    "createId": "ingester",
    "createTime": "2018-11-16T10:44:20.015+01:00",
    "modId": "ingester",
    "modTime": "2018-11-16T10:44:20.015+01:00",
    "complete": false,
    "dataCollectionDatasets": [
    ],
    "datafiles": [
    ],
    "endDate": "2018-11-16T10:44:19.778+01:00",
    "location": "/data/visitor/ls2841/1d16a/fungus_per_wt/fungus_per_wt_100nm_overview_nfp",
    "name": "fungus_per_wt_100nm_overview_nfp",
    "parameters": [
      {
        "id": 150396963,
        "createId": "ingester",
        "createTime": "2018-11-16T10:44:20.640+01:00",
        "modId": "ingester",
        "modTime": "2018-11-16T10:44:20.640+01:00",
        "stringValue": "20",
        "type": {
          "id": 350,
          "createId": "root",
          "createTime": "2014-06-18T15:00:45.996+02:00",
          "modId": "255362",
          "modTime": "2017-11-22T15:11:26.047+01:00",
          "applicableToDataCollection": false,
          "applicableToDatafile": false,
          "applicableToDataset": true,
          "applicableToInvestigation": false,
          "applicableToSample": false,
          "dataCollectionParameters": [
          ],
          "datafileParameters": [
          ],
          "datasetParameters": [
          ],
          "description": "Dark images ",
          "enforced": false,
          "investigationParameters": [
          ],
          "name": "TOMO_darkN",
          "permissibleStringValue": [
          ],
          "sampleParameters": [
          ],
          "units": "None",
          "valueType": "STRING",
          "verified": true
        }
      },
      {
        "id": 150396948,
        "createId": "ingester",
        "createTime": "2018-11-16T10:44:20.590+01:00",
        "modId": "ingester",
        "modTime": "2018-11-16T10:44:20.590+01:00",
        "stringValue": "0.0",
        "type": {
          "id": 351,
          "createId": "root",
          "createTime": "2014-06-18T15:00:46.013+02:00",
          "modId": "255362",
          "modTime": "2017-11-22T15:06:42.624+01:00",
          "applicableToDataCollection": false,

```

EntityManager

[https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT dataset FROM Dataset dataset JOIN dataset.parameters parameters JOIN parameters.type parameterType JOIN dataset.investigation investigation where dataset.id = '150396772' include dataset.parameters.type](https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT%20dataset%20FROM%20Dataset%20dataset%20JOIN%20dataset.parameters%20parameters%20JOIN%20parameters.type%20parameterType%20JOIN%20dataset.investigation%20investigation%20where%20dataset.id%20=%20'150396772'%20include%20dataset.parameters.type)

Data size: 64K

[https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT dataset.id, dataset.name, dataset.startDate, dataset.endDate, investigation.name, parameterType.name, parameter.stringValue, parameter.numericValue FROM DatasetParameter as parameter JOIN parameter.dataset dataset JOIN parameter.type parameterType JOIN dataset.investigation investigation where dataset.id = '150396772' include dataset.parameters.type](https://icat.esrf.fr/icat/entityManager?sessionId=603ebf8f-efeb-46f3-9c08-d00e20337a05&query=SELECT%20dataset.id,%20dataset.name,%20dataset.startDate,%20dataset.endDate,%20investigation.name,%20parameterType.name,%20parameter.stringValue,%20parameter.numericValue%20FROM%20DatasetParameter%20as%20parameter%20JOIN%20parameter.dataset%20dataset%20JOIN%20parameter.type%20parameterType%20JOIN%20dataset.investigation%20investigation%20where%20dataset.id%20=%20'150396772'%20include%20dataset.parameters.type)

Data size: 19K

```
[
  [
    150396772,
    "fungus_per_wt_100nm_overview_nfp",
    "2018-11-16T10:28:24.892+01:00",
    "2018-11-16T10:44:19.778+01:00",
    "LS-2841",
    "TOMO_darkN",
    "20",
    null
  ],
  [
    150396772,
    "fungus_per_wt_100nm_overview_nfp",
    "2018-11-16T10:28:24.892+01:00",
    "2018-11-16T10:44:19.778+01:00",
    "LS-2841",
    "TOMO_yStep",
    "0.0",
    null
  ],
  [
    150396772,
    "fungus_per_wt_100nm_overview_nfp",
    "2018-11-16T10:28:24.892+01:00",
    "2018-11-16T10:44:19.778+01:00",
    "LS-2841",
    "scanType",
    "holonfpscan",
    null
  ],
]
```

API

<https://icatplus.esrf.fr/investigation/{investigationId}/datasets?sessionId=603ebf8f-efeb-46f3-9c0>

Data size: 9.9K

```
{
  "investigation": {
    "name": "LS-2841",
    "id": 150269629
  },
  "id": 150396772,
  "name": "fungus_per_wt_100nm_overview_nfp",
  "startDate": "2018-11-16T10:28:24.892+01:00",
  "endDate": "2018-11-16T10:44:19.778+01:00",
  "location": "/data/visitor/ls2841/id16a/fungus_per_wt/fungus_per_w",
  "parameters": [
    {
      "name": "TOMO_darkN",
      "value": "20"
    },
    {
      "name": "TOMO_yStep",
      "value": "0.0"
    },
    {
      "name": "scanType",
      "value": "holonfpscan"
    },
    {
      "name": "TOMO_zStep",
      "value": "-2.7999999999999998"
    },
    {
      "name": "InstrumentSlitPrimary_vertical_offset",
      "value": "-0.580"
    },
    {
      "name": "InstrumentSlits_name",
      "value": "hpcshsskbs"
    },
    {
      "name": "InstrumentSlitSecondary_name",
      "value": "hpss"
    },
    {
      "name": "InstrumentSlitPrimary_horizontal_gap",
      "value": "2.000"
    },
    {
      "name": "InstrumentSource_mode",
      "value": "7/8multibunch"
    }
  ]
}
```


Search

Name	Beamline	Title	Datasets	Files	Date	
MX-1983	BM29	ISMB application, Biological Sciences, Birkbeck; Structural and Molecular Biology, School of Pharmacy & NPP, UCL; Biological and Chemical Sc	85 (61 GB)	58195	21/11/2018	DOI 10.1515/ESRF-ES-155404944
MX-2077	ID30B	Structural Biology of Macromolecules and Macromolecular complexes from India	283 (261 GB)	224800	21/11/2018	DOI 10.1515/ESRF-ES-155242587
MX-1995	ID30A1	Structural studies of various proteins in Uppsala and Stockholm	499 (323 GB)	141251	21/11/2018	DOI 10.1515/ESRF-ES-155240127
IH-LS-3195	ID17	In-vivo irradiation of rats	2 (22 KB)	2	21/11/2018	DOI 10.1515/ESRF-ES-155216741
MX-2079	ID23-1	Russian BAG for Xtallography and BioSAXS	362 (27 GB)	6372	21/11/2018	DOI 10.1515/ESRF-ES-155205493
IH-MA-63	ID01	Bragg Ptychography of Tungsten Armour			21/11/2018	DOI 10.1515/ESRF-ES-155205381
EV-349	ID16A	Visualization of soil microbe promoted selenium accumulation in young Arabidopsis roots with high-resolution 3D cryo-fluorescence imaging	263 (133 GB)	34435	20/11/2018	DOI 10.1515/ESRF-ES-155204753
			41 (1.7 GB)	7848	19/11/2018	
MX-2087	ID29	Cambridge MRC Block allocation	102 (1 GB)	368	19/11/2018	DOI 10.1515/ESRF-ES-154202231
MX-2141	BM29	To understand the antimicrobial activity of the salivary protein Histatin 5	307 (11 GB)	9506	19/11/2018	DOI 10.1515/ESRF-ES-154128138
IH-MA-69	ID01	Coherent diffraction imaging of LNMO batteries, Si reference structures, Co microcrystals and Ge microdisks			18/11/2018	DOI 10.1515/ESRF-ES-152611353
MX-2064	ID30B	Macromolecular Crystallography at South-East Andalusia	171 (1 TB)	203611	17/11/2018	DOI 10.1515/ESRF-ES-151565774
MX-2066	BM29	Paris Rive-Gauche BAG	172 (6 GB)	5332	17/11/2018	DOI 10.1515/ESRF-ES-151542137
MX-2133	CM01	CryoEM studies of human NAD kinase a novel therapeutic target	45 (5 TB)	2650	16/11/2018	DOI 10.1515/ESRF-ES-150860761
MX-2085	ID23-1	Structural Biology BAG - Portugal	366 (58 GB)	10814	16/11/2018	DOI 10.1515/ESRF-ES-150325184
MX-2119	BM29	Monitoring Assembly Kinetics of Large-Scale DNA Origami Assemblies using Small Angle X-Ray Scattering	341 (1.3 GB)	10565	16/11/2018	DOI 10.1515/ESRF-ES-150291534
LS-2841	ID16A	Understanding the recombination-independent mechanism of DNA homology search and recognition	29 (330 GB)	91952	16/11/2018	DOI 10.1515/ESRF-ES-150269629
MX-2072	BM29	Structural studies of proteins and protein complexes involved in regulation, recognition and transport	56 (2 GB)	1736	15/11/2018	DOI 10.1515/ESRF-ES-149283553
BLC-11865	ID21	Commissioning and performance tests of the new ISDD DCM monochromator at ID21	5 (84 MB)	12	15/11/2018	
ME-1499	ID11	3D characterisation of fatigue crack initiation and growth in a cast Al alloy during gigacycle fatigue.	99 (1 TB)	191174	14/11/2018	DOI 10.1515/ESRF-ES-147900374
					14/11/2018	
			42 (39 GB)	5700	14/11/2018	
MX-2082	BM29	Southampton, UCL Royal Free, Exeter, Bristol, Essex Block Allocation Group (BAG)	22 (178 GB)	165409	14/11/2018	DOI 10.1515/ESRF-ES-147124103
MX-2067	ID29	BAG BARCELONA	1 (0 Bytes)	3	14/11/2018	DOI 10.1515/ESRF-ES-147122999
ES-838	BM05	Multi-scale analysis of early hominin fossils from the Turkana Basin.	964 (6 TB)	4205585	13/11/2018	DOI 10.1515/ESRF-ES-147110748

Showing rows 1 to 25 of 1783

25

1 2 3 4 5 6 7 8 9 10 > >>

Investigations EV-355 How does the model green micro-alga Chlamydomonas reinhardtii adapt to cadmium?

Dataset List 47 Logbook

List Details

Search

Definition	Name	Location	Files	Size	Date	Download
SXM		/de/1um	592	408 MB	28/10/2018 12:28	Download
SXM		/de/he	274	180 MB	28/10/2018 11:34	Download
SXM		/de	431	182 MB	28/10/2018 10:16	Download
SXM		/de/ap1	298	214 MB	28/10/2018 08:57	Download
SXM		/de/s	1210	3 GB	27/10/2018 22:05	Download
SXM		/de	242	105 MB	27/10/2018 21:05	Download
SXM		/de	2	52 KB	27/10/2018 21:05	Download
SXM		/de/S_3cell	8281	428 MB	27/10/2018 18:05	Download
SXM		/de/esmaps	297	80 MB	27/10/2018 17:39	Download
SXM		/de/arise	484	107 MB	27/10/2018 17:04	Download
SXM		/de/ec01	2	37 KB	27/10/2018 17:04	Download
SXM		/de/maps2	400	292 MB	27/10/2018 15:08	Download
SXM		/de/e2	271	50 MB	27/10/2018 14:48	Download
SXM		/de/maps	903	245 MB	27/10/2018 12:41	Download
SXM		/de/e	391	185 MB	27/10/2018 11:45	Download
SXM		/de/01	2	52 KB	27/10/2018 11:45	Download
SXM		/de/S_2c1	7660	275 MB	27/10/2018 09:32	Download
SXM		/de/S_1_2c	8227	546 MB	27/10/2018 06:28	Download
SXM		/de/S_3agg	10788	1 GB	26/10/2018 23:19	Download
SXM		/de/S_3c2	7162	256 MB	26/10/2018 20:57	Download
SXM		/de/S_3c1	6904	237 MB	26/10/2018 18:00	Download
SXM		/de/esmaps	878	399 MB	26/10/2018 16:04	Download
SXM		/de/arise	973	371 MB	26/10/2018 13:34	Download
SXM		/de/ec01	2	37 KB	26/10/2018 13:34	Download

[Investigations](#) EV-355 How does the model green micro-alga *Chlamydomonas reinhardtii* adapt to cadmium?

Dataset List 47

[Logbook](#)

List Details

12:28 PM

October 28, 2018

Scanning X-ray Microscope

Summary

[Metadata List](#)

Files 592

[DOI](#)

Name

Definition

SXM

Start

12:28:00 PM

End

3:40:00 PM

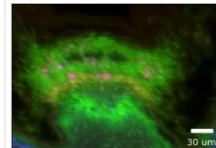
Sample

ble_Cd

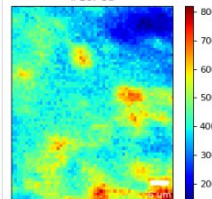
Description

wheat grain 16um thick Cd

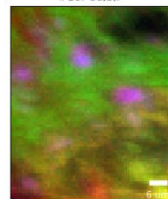
#2: Cd,S,P



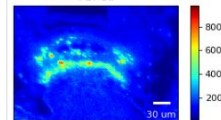
#18: Cd



#18: Cd,S,P



#2: Cd



/data/visitor/ev355/id2 [redacted] um

Download

11:34 AM

October 28, 2018

Scanning X-ray Microscope

Summary

[Metadata List](#)

Files 274

[DOI](#)

Name

Definition

SXM

Start

11:34:00 AM

End

12:28:00 PM

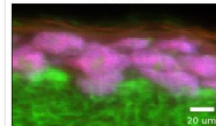
Sample

ble_Cd

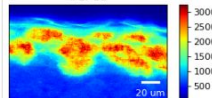
Description

wheat grain 16um thick Cd

#1: Cd,S,P



#1: Cd



New comment

 No tags

Write a title here (optional)



This is a Demo

We can add rich text and images or code:

ls -l

1

Save

Cancel

Sep 24

11:48:55

11:48:55

11:48:55

11:48:55

11:48:55

11:48:55

11:48:55

11-49-55

 Dataset has been collected successfully

Started subscan with value 9

Started subscan with value 8

Started subscan with value 7

Started subscan with value 6

Started subscan with value 5

Started subscan with value 4

Started subseqn with value 3