



| The European Synchrotron

Metadata in ICAT

Samples and Parameters at the ESRF

ICAT F2F meeting - 03/05/2023 - Berlin

Marjolaine Bodin - Data manager - DAU@ESRF

Experiment parameters are stored at the dataset level

Investigation parameters:

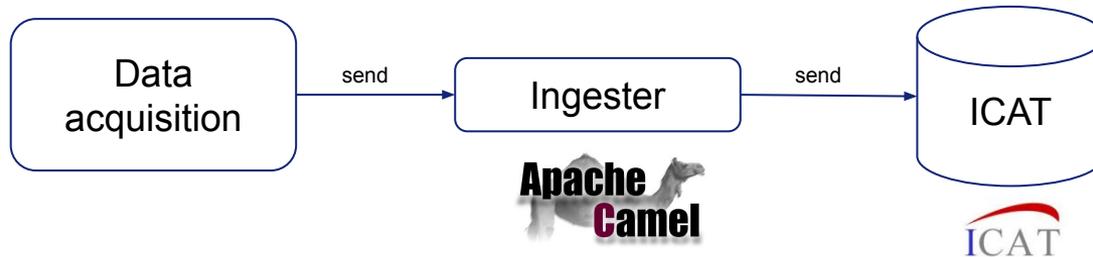
- **Figures: sample count, volume (total, acquisition, processed), file count (total, acquisition, processed)**
- **Identifier used in the User Portal**

Sample parameters:

- **Figures: volume (total, acquisition, processed), file count (total, acquisition, processed)**
- **Information from the User Portal: identifier, acronym, description, safety information**

Dataset parameters

- Stored in the ICAT database: to make them searchable and accessible
- Following Nexus conventions
- XML file describing the mapping between ICAT database and the master file in HDF5 format
- Technique specific information as NXsubentry based keys
- 12 techniques: SAXS, MX, EM, PTYCHO, FLUO, TOMO, MRT, HOLO, WAXS, HTXRPD, SXDM, BCDI



Nexus conventions

```
<?xml version="1.0" encoding="UTF-8"?>
<group NX_class="NXentry" groupName="{entry}">
  <title ESRF_description="Name of the dataset" ESRF_mandatory="Mandatory" NAPIType="NX_CHAR">${datasetName}</title>
  <scanNumber ESRF_description="Scan number" ESRF_mandatory="Mandatory" NAPIType="NX_CHAR">${scanNumber}</scanNumber>
  <proposal ESRF_description="Proposal code" ESRF_mandatory="Mandatory" NAPIType="NX_CHAR">${proposal}</proposal>
  <dataset_type record="final" ESRF_description="Scan type can be 'step_by_step' or 'continuous'" NAPIType="NX_CHAR">${scanType}</dataset_type>
  <folder_path ESRF_description="Scan starting date" ESRF_mandatory="Mandatory" NAPIType="NX_CHAR">${location}</folder_path>
  <start_time ESRF_description="Scan starting date" ESRF_mandatory="Mandatory" NAPIType="NX_DATE_TIME">${startDate}</start_time>
  <end_time ESRF_description="Scan ending date" record="final" ESRF_mandatory="Mandatory" NAPIType="NX_DATE_TIME">${endDate}</end_time>
  <definition record="final" ESRF_description="Techniques used to collect this dataset" NAPIType="NX_CHAR">${definition}</definition>
  <group NX_class="NXsubentry" groupName="SAXS">
    <definition ESRF_description="Technique used to collect this dataset" NAPIType="NX_CHAR">${saxs_definition}</definition>
    <version ESRF_description="Version" NAPIType="NX_CHAR">${saxs_definition.version}</version>
    <directory record="final" ESRF_description="Data collection directory" NAPIType="NX_CHAR">${SAXS_directory}</directory>
    <experimentType record="final" ESRF_description="Type of experiment" NAPIType="NX_CHAR">${SAXS_experimentType}</experimentType>
    <runNumber record="final" ESRF_description="Run number" NAPIType="NX_CHAR">${SAXS_runNumber}</runNumber>
    <prefix record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_prefix}</prefix>
    <maskFile record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_maskFile}</maskFile>
    <numberFrames record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_numberFrames}</numberFrames>
    <timePerFrame record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_timePerFrame}</timePerFrame>
    <concentration record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_concentration}</concentration>
    <comments record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_comments}</comments>
    <code record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_code}</code>
    <detector_distance record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_detector_distance}</detector_distance>
    <waveLength record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_waveLength}</waveLength>
    <pixelSizeX record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_pixelSizeX}</pixelSizeX>
    <pixelSizeY record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_pixelSizeY}</pixelSizeY>
    <beam_center_x record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_beam_center_x}</beam_center_x>
    <beam_center_y record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_beam_center_y}</beam_center_y>
    <normalisation record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_normalisation}</normalisation>
    <diode_currents record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_diode_currents}</diode_currents>
    <acronym record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_acronym}</acronym>
    <transmission record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_transmission}</transmission>
    <storage_temperature record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_storage_temperature}</storage_temperature>
    <exposure_temperature record="final" ESRF_description="" NAPIType="NX_CHAR">${SAXS_exposure_temperature}</exposure_temperature>
```

TechniqueName_entryName

Xml mapping file

<https://gitlab.esrf.fr/icat/hdf5-master-config>

Example 1

Open Data / 10.11111/ESRF-DC-60425398

Dataset List 1

Search

Date	Sample	Dataset	Definition	Files	Size	Processed	Download
16:40 30 Jun 2021	MoSIP3011_pellet	RT		482	34.1 GB		Restore

Summary **Files 482** Metadata List

Search

Name	Value
__elapsedTime	0
__fileCount	482
__volume	36601997282
datasetName	RT
endDate	2021-06-30T16:40:43.266716
sample_name	MoSIP3011_pellet
startDate	2021-06-30T15:53:52.726797
name	RT[2021-06-30 16:40:45]

10

[Restore](#)

ID22 - experiment dataset parameters

No experiment parameters, definition (technique) is empty

Example 3

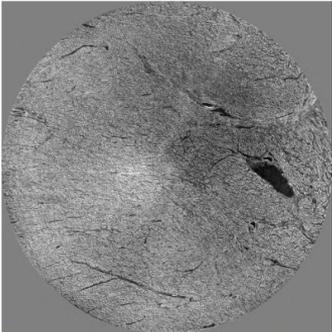
Open Data / 3615151/ESRF-DC-572185639

Dataset List 1

Search

Date	Sample	Dataset	Definition	Files	Size	Processed	Download
13:38 7 Oct 2021	LADAF-2020-27_heart	2.22um_L-vent-muscle	MRtomo	8	38.8 GB		Download

Summary Files 1 Metadata List



The Human Organ Atlas

DOI

Abstract

Title

Users

Patient	
definition	MRtomo
Identifier	LADAF-2020-27
Age (years)	94
Sex	female
Organ	heart
Institute	Laboratoire d'Anatomie des Alpes Francaises
Info	right sylvian and right cerebellar stroke, cognitive disorders of vascular origin, depressive syndrome, atrial fibrillation and hypertensive heart disease, micro-crystalline arthritis (gout), right lung pneumopathy (3 before death), cataract of the left eye, squamous cell carcinoma of the skin (left temporal region)

Sample	
Sample	LADAF-2020-27_heart
Info	complete heart from the body donor program of the Laboratoire d'Anatomie des Alpes Francaise (LADAF)
Preparation	formalin fixed, progressive transfer to ethanol 70%. Vacuum degassing at each transfer to new ethanol bath, mounted with agar crushed gel at 70% ethanol

Scan Parameters	
Instrument	BM05 EBS dipole wiggler 0.85T
SR Current (mA)	200
Exposure Time (s)	0.12
Pixel Size (um)	2.22
Mode (None)	continuous
ScanRadix	2.22um_LADAF-2020.27_heart_LVmus
Step (x,y,z)	--4
Stages (x,y,z)	1,1,8
Projections	6000
rebin	Click to edit
darkn (None)	400
refcon	Click to edit
Acc. Frames Count	1
Detector Distance (mm)	970
Energy (avg) (keV)	84
Scan Geometry	half-acquisition
Scan Range (deg)	360
Pixel (x,y)	2048,2048
Magnification	3
Scintillator	LuAG:Ce 100 um + lead glass meniscus
Sur. Dose Rate (Gy/s)	122
Dose Rate (Gy/s)	35.9
VOI Integ. Dose (kGy)	13
Scan time (min)	12
Series time (h)	1.6

Sensor	
Name	sCMOS PCO edge 4.2 CLHS
Mode	rolling shutter
Size (um)	6.5
Optics Type	Hasselblad tandem optic 100mm/300mm

Processing	
refapproach	reference jar with 70% ethanol, single reference per scan type
Volume X	3775
Volume Y	3775
Volume Z	12891
32to16bitsmin	-0.07
32to16bitsmax	0.08
jp2compratio	10
filters	Mo 0.25mm SIO2 bars 3*4mm diameter
technique	Hierarchical Phase-Contrast Tomography
experimentType	tomography

Human Organ Atlas (processed) dataset parameters

Link between raw and processed datasets:

Processing / Workflow

Add metadata `input_datasets` with the (list of) path(s) of raw data

Ingester

Match path and raw dataset ids (`input_datasetIds`)

Add metadata on raw datasets (`output_datasetIds`, `output_datasets`)

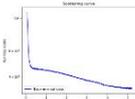
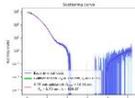
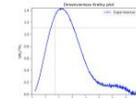
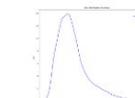
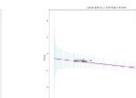
Allows us to display processed data in the data portal:



Processing and workflow information

```

<group NX class="NXsubentry" groupName="SAXS">
  <definition ESRF_description="Technique used to collect this dataset" NAPItype="NX_CHAR">${SAXS_definition}</definition>
  ...
  <transmission record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_transmission}</transmission>
  <storage temperature record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_storage_temperature}</storage temperature>
  <exposure temperature record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_exposure_temperature}</exposure temperature>
  <column_type record="final" ESRF_description="HPLC column type. [ex. Agilent BioSEC 130]" NAPItype="NX_CHAR">${SAXS_column_type}</column_type>
  <flow_rate record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_flow_rate}</flow_rate>
  <hplc_port record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_hplc_port}</hplc_port>
  <sample_type record="final" ESRF_description="It can be buffer or sample" NAPItype="NX_CHAR">${SAXS_sample_type}</sample_type>
  <run_number record="final" ESRF_description="It can be buffer or sample" NAPItype="NX_CHAR">${SAXS_run_number}</run_number>
  <experiment_type record="final" ESRF_description="It the kind of experiment: sample changer or HPLC" NAPItype="NX_CHAR">${SAXS_experiment_type}</experiment_type>
  <guinier_rg record="final" ESRF_description="Guinier radius of giration" NAPItype="NX_CHAR">${SAXS_guinier_rg}</guinier_rg>
  <guinier_points record="final" ESRF_description="Points of the guinier region" NAPItype="NX_CHAR">${SAXS_guinier_points}</guinier_points>
  <guinier_i0 record="final" ESRF_description="Guinier radius of giration" NAPItype="NX_CHAR">${SAXS_guinier_i0}</guinier_i0>
  <rg record="final" ESRF_description="Calculated raiodius of giration. It can be calculated with Gnom or BIFT" NAPItype="NX_CHAR">${SAXS_rg}</rg>
  <total record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_total}</total>
  <d_max record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_d_max}</d_max>
  <porod_volume record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_porod_volume}</porod_volume>
  <porod_MM volume estimation record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_porod_MM volume estimation}</porod_MM volume estimation>
  <frames_averaged record="final" ESRF_description="" NAPItype="NX_CHAR">${SAXS_frames_averaged}</frames_averaged>
  
```

	Sample Changer	Run	Frames		Guinier			BIFT			Porod		MM vol. est.	Scattering	Kratky	Density	Guinier	
			Avq/Total	Time	Rg	Points	IO	Rg	Total	D _{max}	Volume							
05:15:37	buffer_after_ dC 20.0 C 0.0 mg/ml ✓Integrate	# 18	2-7/10	10 s														Download
/data/visitor/mx2484/bm29/20230421/																		
05:14:08	sample_ dC 20.0 C 14.0 mg/ml ✓Integrate ✓Subtract	# 17	5-6/10	1.0 s	3.6±0.2 nm	25-30	831.6±21.3 NA	3.7±0.0 NA	16.4±0.0 NA	205.82 nm ³								Download
/data/visitor/mx2484/bm29/20230421/																		

Experiment Parameters in ICAT

- **Store all metadata at the dataset level (including processing job): is it the right way?**
- **Do we need to have relations between parameters?**
- **Units: what is the best way to handle them?**
- **Search by parameters: performance?**

Current situation



synchronization



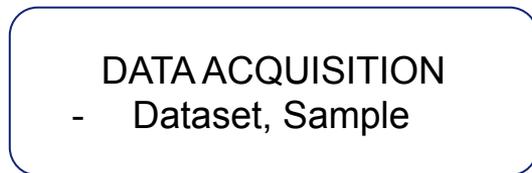
Sample information are stored into `SampleParameter` UP identifier, description, acronym, safety



+ Add an item

Name	Type	Sample	Description
laser	Tool		at 400nm
sample1	Samplesheet	Samplesheet E	

10 ▾



push



BNA
BNA-TL4RIXS23-001
BNA-TL4RIXS23-003
BNA-TL4RIXS23-003b
BNAP-TL4RIXS23-004

SampleSheet

Samples (data acquisition)

Issues

- How to deal with a large number of samples?

Content

Name*

Sample1_01

Sample Sheet

Select sample...

Description (from sample)

Comments

Save Cancel

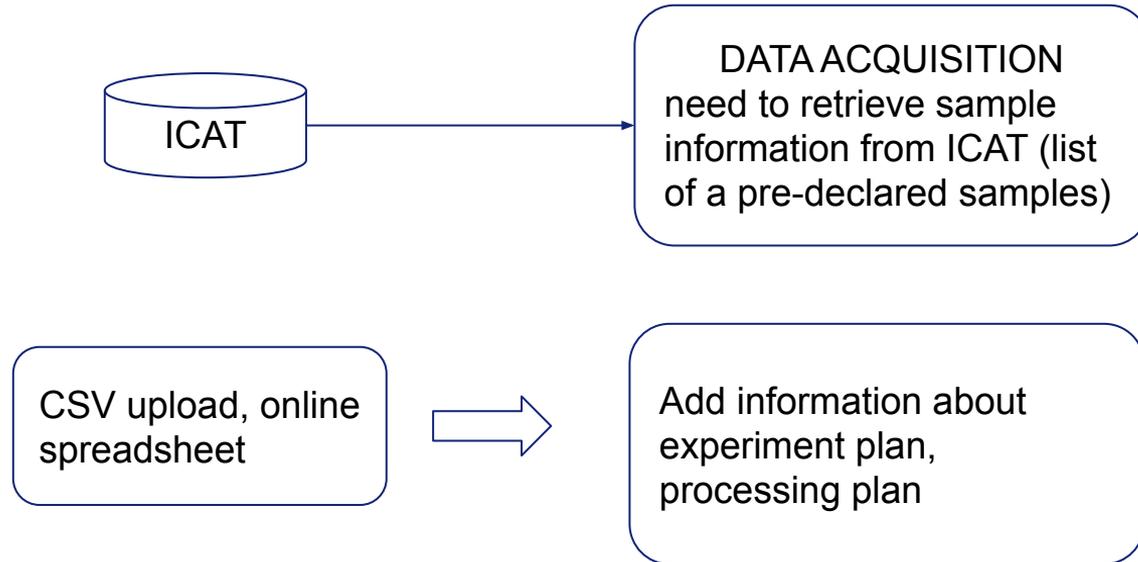


CSV upload, online spreadsheet

Manual input of the sample,
one by one

Issues

- Link with data acquisition sample



Sample in ICAT

- **How to represent sampleSheet vs sample?**
- **SampleType is not used:**
 - Sample has 1 sampleType (does not fit with sampleSheet)
 - SampleType.molecularFormula is mandatory (#239)
- **Need to store experiment plan, processing plan (different from one beamline/technique to another)**