ESRF Development Status

Alex de Maria Antolinos
Software Engineer
Data Manager@Data Analysis Unit
Software Group
ESRF
20.11.1017

Data Policy Short Summary

- Acceptance of the data policy is a **condition** for the **award of beam time**
- ESRF will automatically collect data and metadata for all experiments from all beamlines (including CRGs)
- ESRF is the **custodian** of the **raw data and associated metadata**
- ESRF will store metadata in a metadata catalogue (**ICAT**)
- Experimental team has sole access to the data during the so-called embargo period of 3 years; request to extend the embargo period can be made
- **After the embargo ESRF will make the data "Open Access"**
- **Proprietary data belong by default to the PI and are not archived unless explicitly agreed**

Data Policy Short Summary

- About data and metadata
  - Only keep **data generated** at the ESRF
  - Data must be in a **format** the ESRF can read
  - Data must be traceable and verifiable as coming from the ESRF

- After the embargo the data will be released under the license CC-By-4

Status

http://www.esrf.fr/datapolicy

Beamline	Status	Techniques	Data Policy Implementation Metadata Collection*	Data archiving*	Raw Data in HDF5*	Open access to data*
ID01		KMAP	implemented	in progress		
BM01A						
BM01B						
ID02						
BM02						
ID03						
BM05		Tomography	in progress	in progress		
ID06-LVP						
ID06						
BM08						
ID09						
ID10						
ID11		Tomography	in progress	in progress		
ID12						
ID13						
BM14						
BM15A						
BM15B						
ID16A		Fluo, Tomo	implemented	in progress	implemented	
ID16B		Tomo	in progress	in progress		
ID17		MRT, Tomography	implemented	in progress		
ID18						
ID19		Tomography	in progress	in progress		
ID20			in progress	in progress		
ID21		Microscopy	implemented	in progress		
ID22						
ID23-1		MX	implemented	in progress		
ID23-2		MX	implemented	in progress		
BM23						
ID24						
BM25A						
BM25B						
ID26						
BM26A						
BM26B						
ID27						
ID28						
BM28						
ID29		MX	implemented	in progress		
BM29		BIOSAXS	implemented	in progress		
ID30A-1		MX	implemented	in progress		
ID30A-3		MX	implemented	in progress	implemented	
ID30B		MX	implemented	in progress		
BM30A						
BM30B						
ID31						
ID32			in progress	in progress		
BM32						
CryoEM						

***Techniques covered** - lists which techniques are currently concerned by the Data Policy

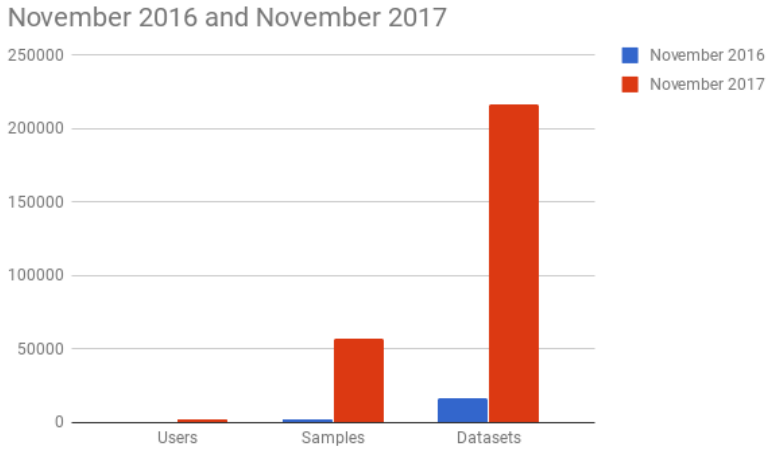
***Metadata collection** - status of metadata collection and storage in metadata catalogue for the listed techniques (orange = planned / in progress, green = implemented, grey = not planned / implemented yet)

***Data archiving** - status of long term archiving in tape library (green = data are being archived for 10 years in tape archive)

***Open access of data** - indicates if data is open access (green) or still under embargo (red), or no data archived (grey)

Status

One year ago….. And now



x8 x27 x12

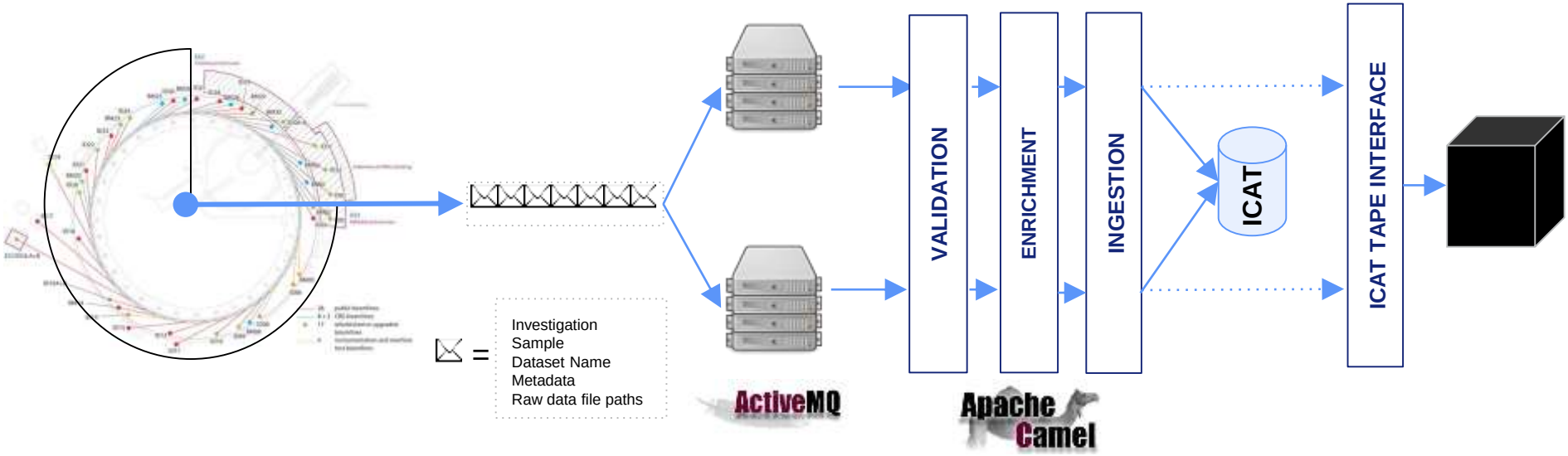
	October 2016	November 2017
Users	272	2333
Samples	2065	57363
Datasets	16980	216570
Parameters	472000	5973581
Ratio (Parameters/dataset)	27.79740872	27.58267996

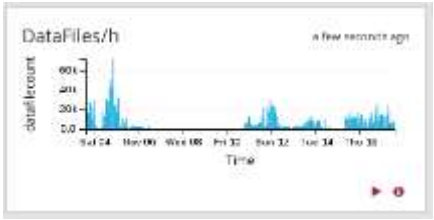
Architecture Overview

Data: from Beamline to Tape

Architecture

https://icat.esrf.fr





PRODUCERS

CONSUMERS

Metadata

How and where metadata is stored?

HDF5 + Nexus + ICAT

https://icat.esrf.fr

NeXus

ICAT

HDF5 + Nexus + ICAT

https://icat.esrf.fr

- HDF5 as a mirror of ICAT on the local beamline file system
- Following the NEXUS convention

```
−<group NX_class="NXentry" groupName="${entry}">
    <title ESRF_description="Name of the dataset" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${scanName}</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 ESRF_description="Scan type can be 'step_by_step' or 'continuous'&#xA;&#x9;&#x9;" 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 ESRF_description="Techniques used to collect this dataset" NAPItype="NX_CHAR">${definition}</definition>
  +<group NX_class="NXsubentry" groupName="SAXS"></group>
  +<group NX_class="NXsubentry" groupName="MX"></group>
  +<group NX_class="NXsubentry" groupName="PTYCHO"></group>
  +<group NX_class="NXsubentry" groupName="FLUO"></group>
  +<group NX_class="NXsubentry" groupName="TOMO"></group>
  +<group NX_class="NXsubentry" groupName="MRT"></group>
  +<group NX_class="NXsubentry" groupName="HOLO"></group>
  +<group NX_class="NXsubentry" groupName="WAXS"></group>
  +<group NX_class="NXsample" groupName="sample"></group>
  +<group NX_class="NXinstrument" groupName="instrument"></group>
  +<group NX_class="NXnote" groupName="notes"></group>
  </group>
```

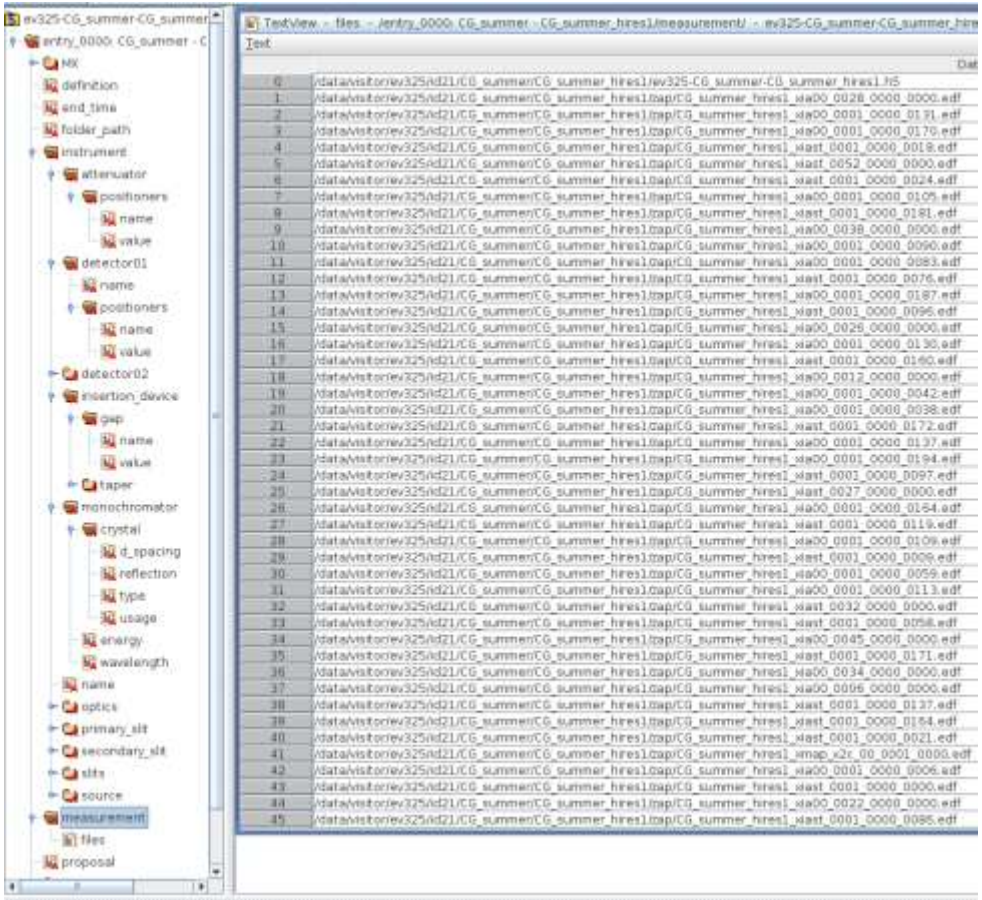
HDF5 + Nexus + ICAT

https://icat.esrf.fr

```
−<group NX_class="NXentry" groupName="${entry}">
    <title ESRF_description="Name of the dataset" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${scanName}</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 ESRF_description="Scan type can be 'step_by_step' or 'continuous'&#xA;&#x9;&#x9;" 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 ESRF_description="Techniques used to collect this dataset" NAPItype="NX_CHAR">${definition}</definition>
  +<group NX_class="NXsubentry" groupName="SAXS"></group>
  +<group NX_class="NXsubentry" groupName="MX"></group>
  +<group NX_class="NXsubentry" groupName="PTYCHO"></group>
  +<group NX_class="NXsubentry" groupName="FLUO"></group>
  +<group NX_class="NXsubentry" groupName="TOMO"></group>
  +<group NX_class="NXsubentry" groupName="MRT"></group>
  +<group NX_class="NXsubentry" groupName="HOLO"></group>
  +<group NX_class="NXsubentry" groupName="WAXS"></group>
  −<group NX_class="NXsample" groupName="sample">
      <name ESRF_description="Name of the sample" ESRF_mandatory="Mandatory" NAPItype="NX_CHAR">${Sample_name}</name>
      <description ESRF_description="Description of the sample" NAPItype="NX_CHAR">${Sample_description}</description>
      <chemical_formula ESRF_description="Chemical formula of the sample" NAPItype="NX_CHAR">${Sample_chemical_formula}</chemical_formula>
    −<group NX_class="NXpositioner" groupName="positioners">
        <name NAPItype="NX_CHAR">${SamplePositioners_name}</name>
        <value NAPItype="NX_CHAR">${SamplePositioners_value}</value>
      </group>
    −<group NX_class="NXenvironment" groupName="environment">
      −<group NX_class="NXsensor" groupName="sensors" ESRF_description="Parameters for controlling external conditions">
          <name NAPItype="NX_CHAR">${SampleEnvironmentSensors_name}</name>
          <value NAPItype="NX_CHAR">${SampleEnvironmentSensors_value}</value>
        </group>
      </group>
    </group>
  +<group NX_class="NXinstrument" groupName="instrument"></group>
  +<group NX_class="NXnote" groupName="notes"></group>
  </group>
```

HDF5 + Nexus + ICAT

https://icat.esrf.fr



- Parameter type are named as with Nexus convention style

Name	Value
InstrumentAttenuatorPositioners_name	"att1 att2 att3 att4 att5 ic0"
InstrumentAttenuatorPositioners_value	"C_v_1.15 Al_0.28 Al_1.24 Cu_0.35 Cu_0.69 Unknown"
InstrumentInsertionDevice_gap_name	W125U W150M
InstrumentInsertionDevice_gap_value	199.999 24.800
InstrumentPositioners_name	" trash tth th chi phi kth kap kphi ky saz sax say muy mutilt muz murot slhtx slhg slvtx colly pitch yaw roll sly slth slz slchi slvz fry frz focus rotc opt1 slit mrtvo mrtho mrtvg mrthg videohz videovz k0 sigap sloff waterx watery waterz scsfy scsyaw scsz scsy c1th c2th srot w125u colth coltx xc collth collz n
InstrumentPositioners_value	" 1 0 0 0 0 0 0 0 16.5 42.455 17 0 0.2 0 0 0 0 0.4 -2.0413 10 0 0 0 -0.292 0 -50 0 76.164 293 21 -0.215221 -4.6575 -3.33333e-06 0 0 4 5 24.8 0 100 0.53005 30.536 37.9996 0 0 2 0.1 -49.95 25.66 0 0 0 0.12 0 -200 1.896 1.909 -200.29 199.999 0.949995 -0.0205 0.000249237 -1.07368 16.885 0

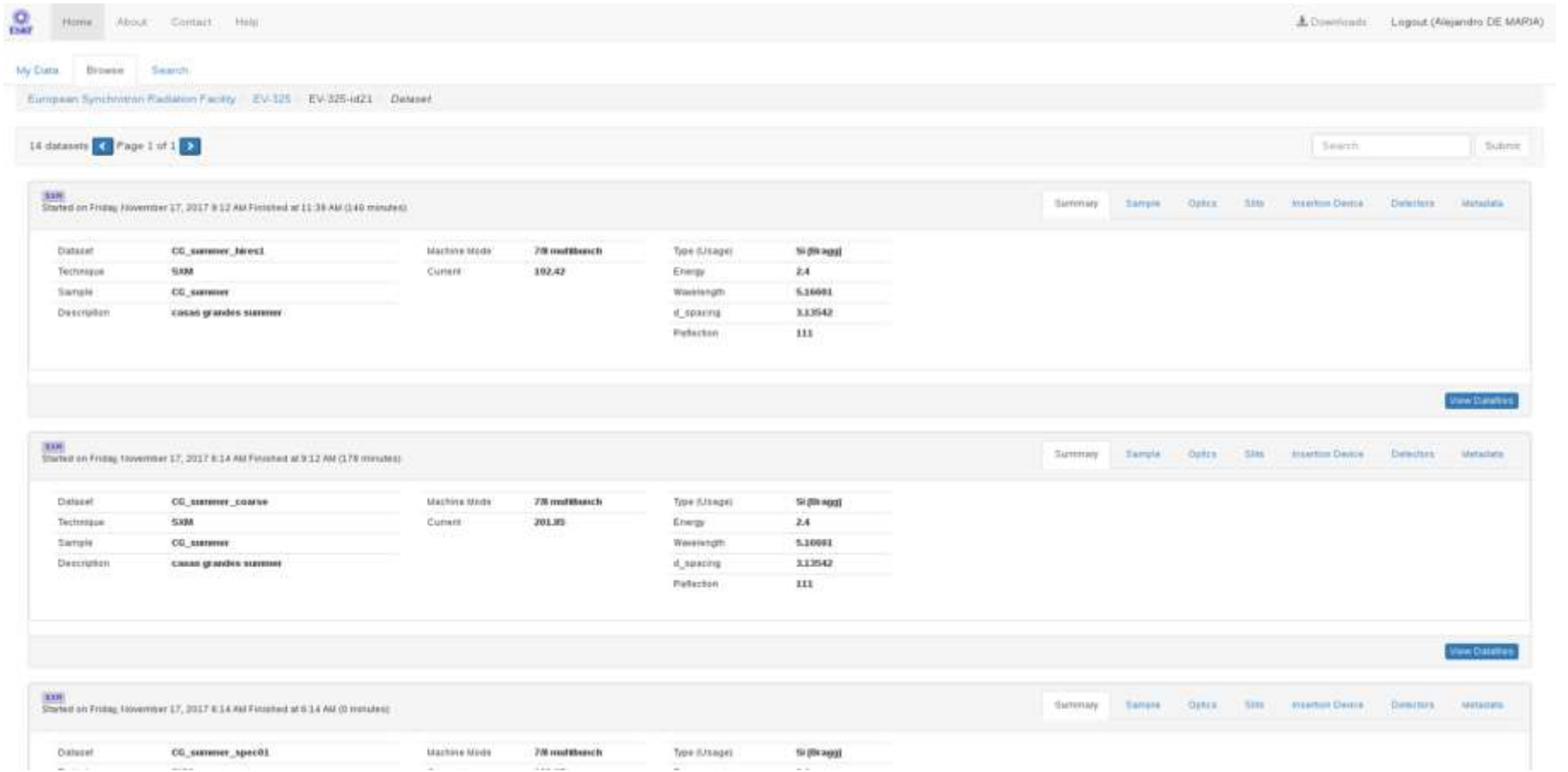
Problems (?):

- Values as arrays on the database
- Topcat can not display parameter type names

TopCAT Developments

Metadata plugin

TopCAT


Home
About
Contact
Help
Downloads
Logout (Alejandro DE MARIA)
My Data
Browse
Search
European Synchrotron Radiation Facility
EV-325
EV-325-id21
Dataset
14 datasets
Page 1 of 1
Search
Submit
SXM
Started on Friday, November 17, 2017 9:12 AM Finished at 11:38 AM (140 minutes)
Summary
Sample
Optics
Slits
Insertion Device
Detectors
Metadata
Dataset
CG_summer_hires1
Technique
SXM
Sample
CG_summer
Description
casas grandes summer
Machine Mode
7/8 multibunch
Current
192.42
Type (Usage)
Si (Bragg)
Energy
2.4
Wavelength
5.16691
d_spacing
3.13542
Reflection
111
View Datafiles
SXM
Started on Friday, November 17, 2017 6:14 AM Finished at 9:12 AM (178 minutes)
Summary
Sample
Optics
Slits
Insertion Device
Detectors
Metadata
Dataset
CG_summer_coarse
Technique
SXM
Sample
CG_summer
Description
casas grandes summer
Machine Mode
7/8 multibunch
Current
201.85
Type (Usage)
Si (Bragg)
Energy
2.4
Wavelength
5.16691
d_spacing
3.13542
Reflection
111
View Datafiles
SXM
Started on Friday, November 17, 2017 6:14 AM Finished at 6:14 AM (0 minutes)
Summary
Sample
Optics
Slits
Insertion Device
Detectors
Metadata
Dataset
CG_summer_spec01
Machine Mode
7/8 multibunch
Type (Usage)
Si (Bragg)

TopCAT

Home | About | Contact | Help | Downloads | Logout (Alejandro DE MARIA)

My Data | Browse | Search

European Synchrotron Radiation Facility / MX-1932 / MX-1932-id29 / *Dataset*

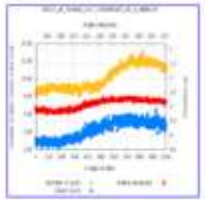
241 datasets < Page 1 of 4 >

Search | Submit

Started on Saturday, October 21, 2017 10:44 AM Finished at 10:45 AM (0 minutes)

Summary | Positioners

Dataset	hAOX_wt_formate_pH7_100mMNaCl_w1_4_2132387	Machine Mode	7/8 multibunch	Flux	6.94e+11		
Technique	OSC	Oscillation Overlap	0.0	Current	192.79		
# Images	1600	Aperture	3	Resolution	2.50005226645		
Template	hAOX_wt_formate_pH7_100mMNaCl_w1_4_####.cbf	Wavelength	0.976251275026	Transmission	15.9579508094		
Detector Distance	587.50625	Mx_beamShape	ellipse				
Beam Size @Sample	0.05, 0.03						

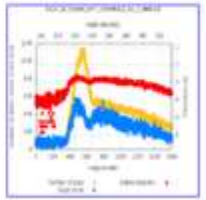


View Datafiles

Started on Saturday, October 21, 2017 10:34 AM Finished at 10:35 AM (1 minutes)

Summary | Positioners

Dataset	hAOX_wt_formate_pH7_100mMNaCl_w1_3_2132381	Machine Mode	7/8 multibunch	Flux	4.26e+11		
Technique	OSC	Oscillation Overlap	0.0	Current	193.66		
# Images	1600	Aperture	3	Resolution	2.00499109738		
Template	hAOX_wt_formate_pH7_100mMNaCl_w1_3_####.cbf	Wavelength	0.976251275026	Transmission	9.93571351576		
Detector Distance	575.00625	Mx_beamShape	ellipse				
Beam Size @Sample	0.05, 0.03						

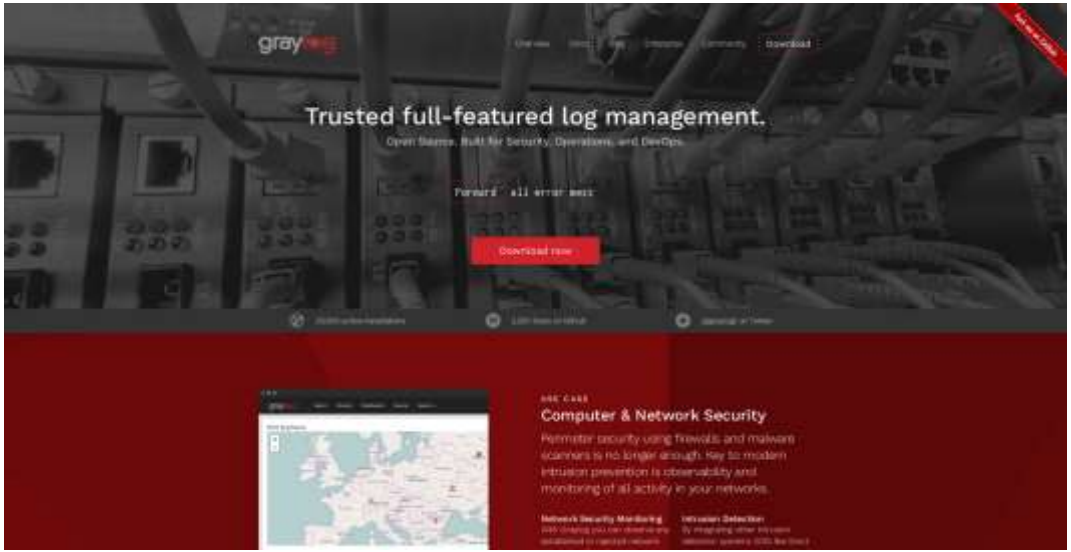


Graylog

Central logging system

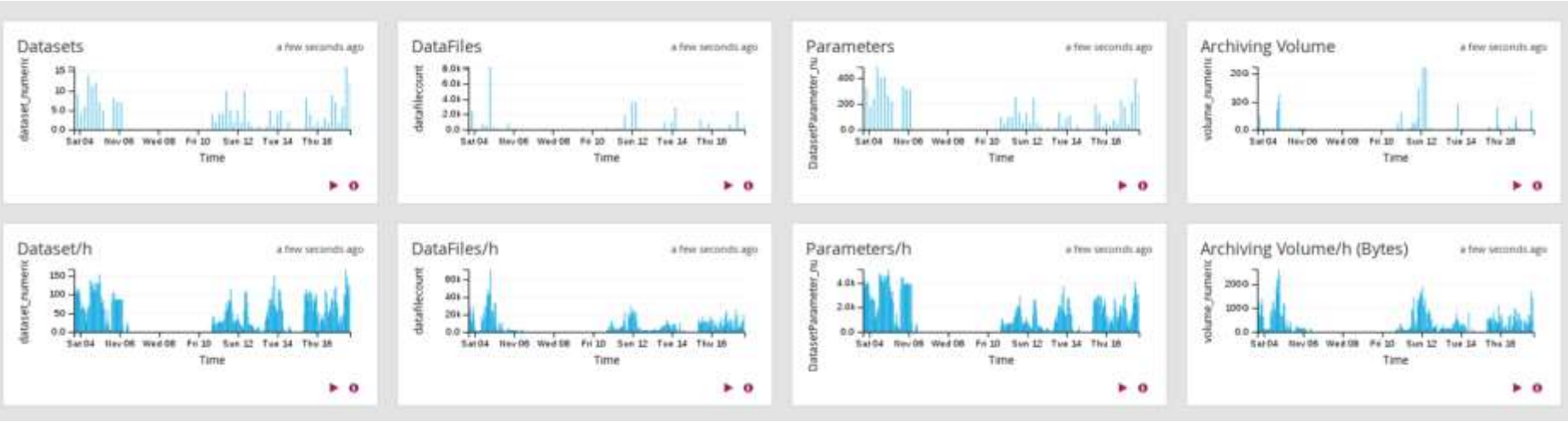
GrayLog

- Centralize all logs
- Open Source
- Monitor and alert
- Stats
- Implementation in Java for all major logging frameworks: log4j, log4j2, java.util.logging, logback, JBossAS7 and WildFly 8- ...



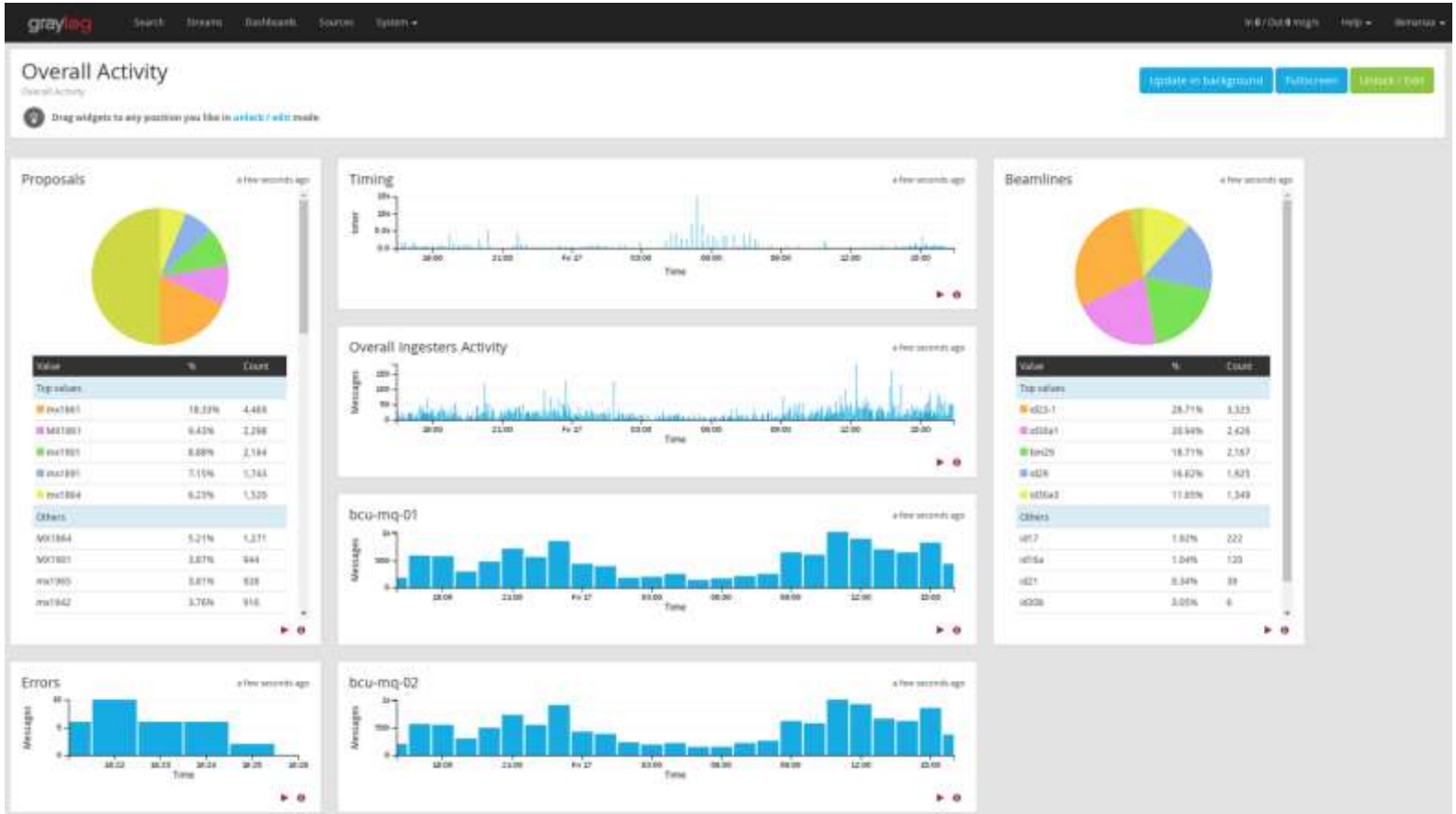
Graylog

http://www.esrf.fr/datapolicy



Graylog

http://www.esrf.fr/datapolicy



Next Developments

- **DOIs**
- **More protocols to download files**
- **E-logbook and its integration with ICAT**
- **Improve/evolve TopCAT**
  - Searchable
  - More attractive

Challenges

- **Attract users with TopCAT**
- **Keep pushing the implementation on the beamlines 10/year**

Thanks!