

Scientific Data Management (SDM)

Mikel Eguiraun - MAX IV SW Team

Data Policy

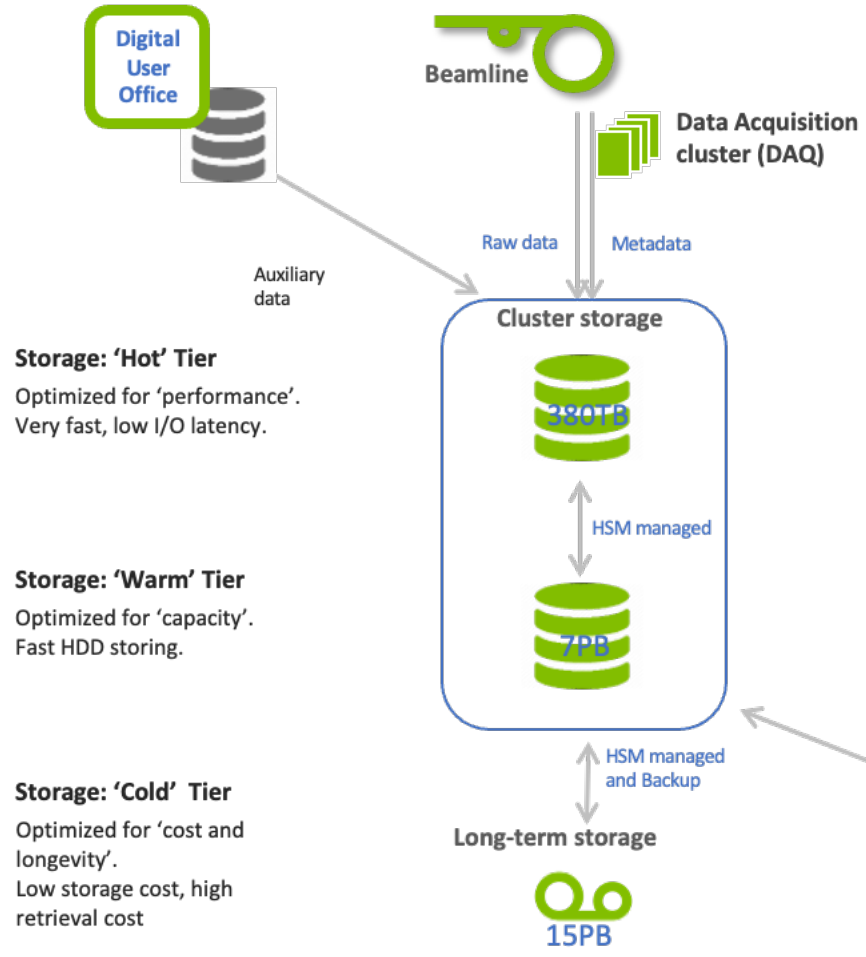
- **MAX IV claim no ownership** to experimental data (raw data and associated metadata) -> MAX IV acts as the custodian of experimental data.
- Metadata automatically captured by instruments is either curated within the raw data files or within a associated online catalogue.
- Experimental data is **preserved for a minimal of seven (7) years** and will be **read-only** for the duration of the life time at MAX IV.
- Raw data and associated metadata is accessible and searchable via **online catalogue**.

Data Storage

Research lifecycle



Data freshness

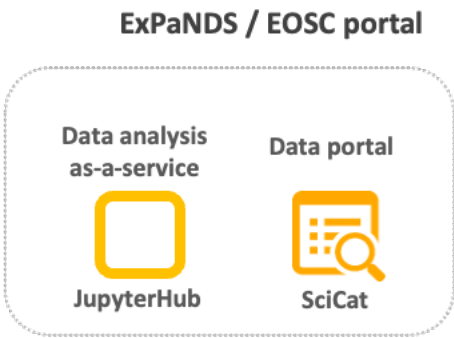


Storage: 'Hot' Tier
Optimized for 'performance'.
Very fast, low I/O latency.

Storage: 'Warm' Tier
Optimized for 'capacity'.
Fast HDD storing.

Storage: 'Cold' Tier
Optimized for 'cost and longevity'.
Low storage cost, high retrieval cost

	2018	2023
User visits	370	1300
Total data size	1.6 PB	2.5 PB
Data/year	~0.2 PB	~1 PB



Data Access

- Globus Connect
- Sftp
- Rclone (targeted proposals, for cloud containers)
- 2 x 10Gbit/s to SUNET, bypassing LU packet inspection
- R/O NFS at LUNARC
- Additional services in development

Ownership

`/data/(user-type)/(beamline)/(proposal)/(visit)/raw`

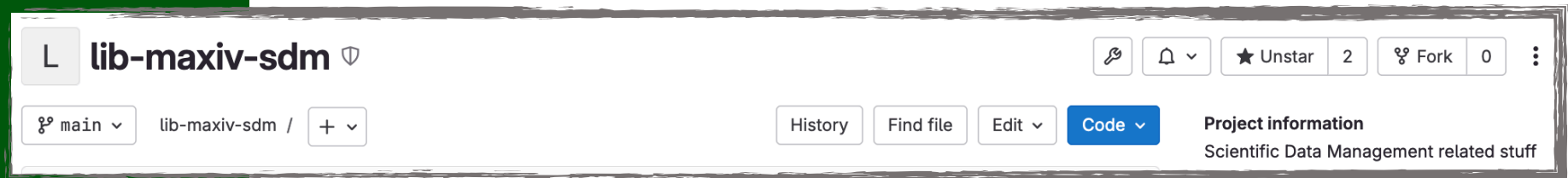
- The (user-type) includes visitors, staff and proprietary
- The (visit) should begin with the date when the first shift starts
- In addition to /raw there will be a /process folder at the same level
 - Data saving services runs as the corresponding service account

	/data/visitor	/beamline	/proposal	/visit/raw
owner	root	service account	service account	service account
group	root	beamline group	proposal group	proposal group
Permissions	2755	2775 or 2755	2750	2770
	static	static	dynamically created	dynamically created

`/data/visitors/biomax/20170251/20240121/raw`

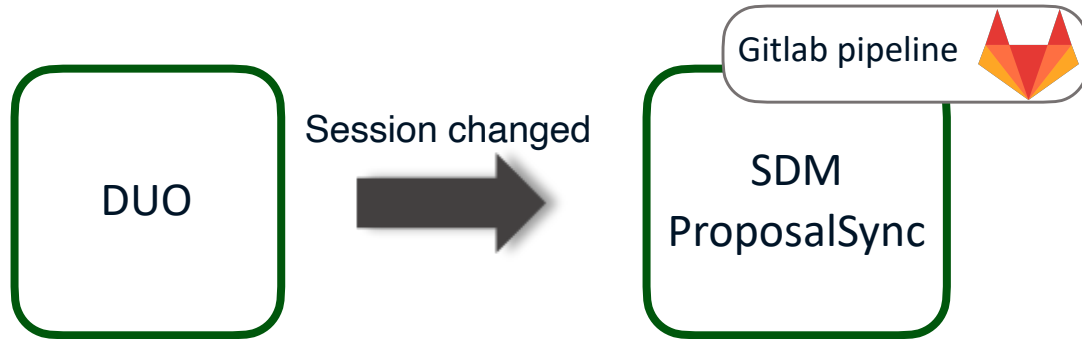
SDM

- SDM is (also) a python library
- Links User Office with experimental control system
- Ensures that the data goes to the right place, with the right permissions



The screenshot shows the GitHub interface for the repository `lib-maxiv-sdm`. The repository name is displayed with a shield icon indicating it is a library. On the right side, there are icons for search, notifications, unstar (2), and fork (0). Below the repository name, there are buttons for `main`, `lib-maxiv-sdm /`, and `+`. The navigation bar includes `History`, `Find file`, `Edit`, and `Code` (highlighted in blue). On the right, there is a `Project information` section with the text `Scientific Data Management related stuff`.

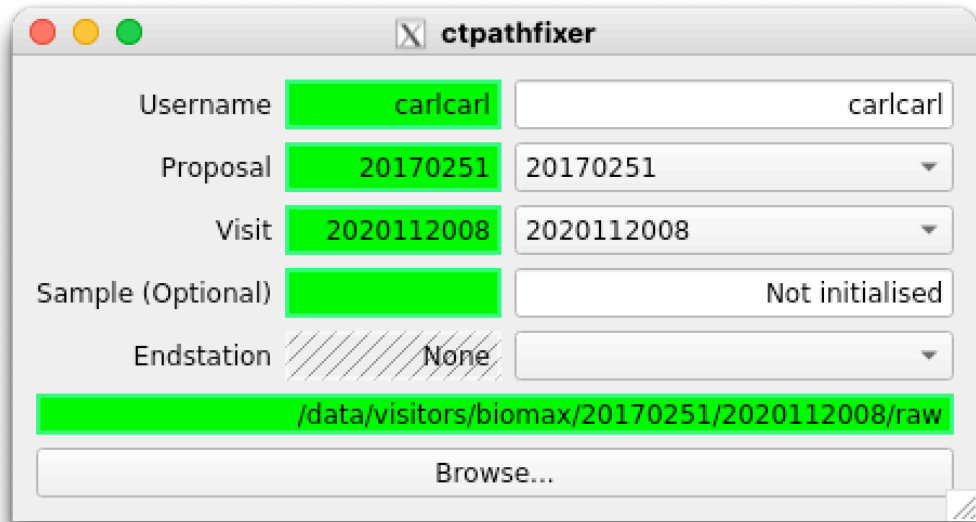
DUO-SDM before experiment



- check and create proposal group in AD
- for all the proposal users in DUO:
 - Sync with the AD proposal group members (DUO is the truth)
 - Move the user to DUOACTIVE if not there already
 - Add to Visitors AD group, add extra_attributes in AD, enable, create_home
 - Add the user to the proposal group
- add the beamline group to the proposal group
- associate extra resources to the proposal group (e.g. presto, vpn)

DUO-SDM during experiment

- PathFixer is a tango device



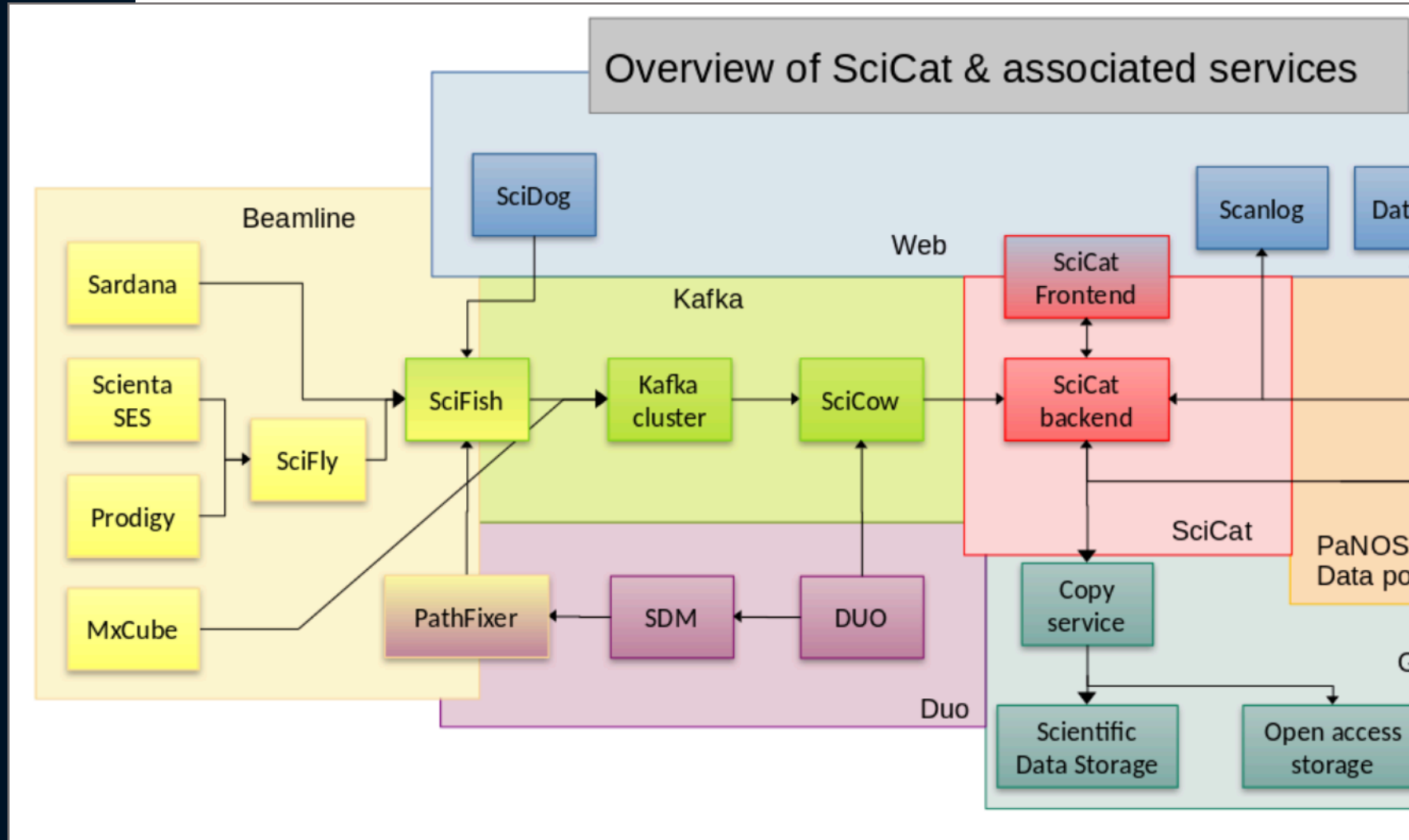
The screenshot shows a window titled "ctpathfixer" with the following fields:

Username	carlcarl	carlcarl
Proposal	20170251	20170251
Visit	2020112008	2020112008
Sample (Optional)		Not initialised
Endstation	None	None
/data/visitors/biomax/20170251/2020112008/raw		
Browse...		

- Carlcarl is the user at the beamline
- Select a proposal from the available ones (typically accepted proposals)
- Select a visit
- Optional extra sample folder
- Path is created on disk, with correct permissions for this user/proposal
- This new path is propagated to the control system (e.g. Sardana)
- (Tango device runs as service account, so full data access)

NO SESSION
NO data

What about metadata?



Scientific Metadata Configuration

Show configuration: All ▾

Group	Name	Device	Attribute	Unit	Mapping	Edit	Delete
	R1 ring current	expchan/ring_current_ctrl/1	value	mA			
	Branch selected	b107a-o03/opt/branch-selector-01	a_selected		True -> A False -> B		
	A branch end station	b107a-0a05/opt/endstation-selector-01	ea01_selected		True -> EA01 False -> EA02		
	B branch end station	b107a-ob05/opt/endstation-selector-01	eb01_selected		True -> EB01 False -> EB02		
	Pressure EA01	b107a-d100730cab08/vac/vgcul-02	pressure_1	mbar			
	Pressure EB01	B107A-EB01/VAC/VGCP-01	pressure	mbar			
	Pressure EB02	B107A-EB02/VAC/VGCP-01	pressure	mbar			
Exit slit positions	a slit1 h	b107a-0a04/opt/slit-01-h	position	um			
Exit slit positions	a slit2 v real	pm/real_a_vslit_ctrl_0/1	position	um			
Exit slit positions	b slit1 h	b107a-ob04/opt/slit-01-h	position	um			
Exit slit positions	b slit2 v real	pm/real_b_vslit_ctrl_0/1	position	um			
Manipulator motor position EA01	mp01 x	motor/simstep_ctrl_0/1	position				
Manipulator motor	mp01 y	motor/simstep_ctrl_0/2	position				

Add Metadata

Group

Name

Device


Attribute

Unit

Add new mapping

Submit

SCANLOG



SciCat Scanlog - BioMAX

Logged in as *mikegu*
Log out

Showing all scans from selected Proposal ID

20170251 (5926)

ScanID
 Start time
 PI
 Proposal
 Name
 Scan
 Detectors
 Data path
 SciCat
 Comment
 Upload Image
 Images
 Data Quality Metrics

x | v

<input type="checkbox"/>	ScanID	Start time ▼	PI	Proposal	Name	Scan	Detectors	Data path	SciCat	Comment
<input type="checkbox"/>	undefined	2024-09-26 10:53:38.907	Ana.Gonzalez@maxiv.lu.se	20170251	Bo-EA1_alberto_1_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240926/raw/d/bo/bo-EA1_alberto	Link	
<input type="checkbox"/>	undefined	2024-09-08 18:03:08.920	Ana.Gonzalez@maxiv.lu.se	20170251	Air-airaft_hv_reset_2_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240908/raw/air/air-air	Link	
<input type="checkbox"/>	undefined	2024-09-04 09:49:46.052	Ana.Gonzalez@maxiv.lu.se	20170251	Plastic-s1_4_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240904/raw/plastic/plastic-s1	Link	
<input type="checkbox"/>	undefined	2024-09-04 09:45:37.196	Ana.Gonzalez@maxiv.lu.se	20170251	Air_1_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240904/raw/plastic/plastic-s1	Link	
<input type="checkbox"/>	undefined	2024-09-04 09:37:57.710	Ana.Gonzalez@maxiv.lu.se	20170251	Ref-air_1_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240904/raw/plastic/plastic-s1	Link	
<input type="checkbox"/>	undefined	2024-09-04 09:34:32.025	Ana.Gonzalez@maxiv.lu.se	20170251	Ref-plastic-s1_4_master.h5	undefined	undefined	/data/visitors/biomax/20170251/20240904/raw/plastic/plastic-s1	Link	

Rows per page: 10 ▼ 1-10 of 5926 < < > >

Add ScanID range

Add ScanIDs

Export no data selected

Title:

Logbook:


Create a new entry in **no logbook selected**

Create group no data selected

Select...

Group

* can't find the group? Type in and create new!



Status

	Beamline	Status	Host	Control system
1	BioMAX	Production	MxCube b-v-biomax-ec-2	MxCube Sardana
2	CoSAXS	Production	b-v-cosaxs-ec-0	Sardana
3	FlexPES	Production	flexpes-scifly-daq.daq.maxiv.lu.se b-v-flexpes-ec-0	Scienta Sardana
4	DanMAX	Production	b-v-danmax-ec-0	Sardana
5	MAXPEEM	Production	maxpeem-scifly-daq.daq.maxiv.lu.se b-v-maxpeem-ec-0	Uview Sardana
6	Veritas	Production	b-v-veritas-ec-0	Sardana
7	FinESTBeAMS	Production	b-v-finest-ec-0	Sardana
8	Balder	Production	b-v-balder-ec-0	Sardana
9	NanoMAX	Production	b-nanomax-controlroom-cc-3	Contrast
10	ForMAX	Production	b-v-formax-ec-2	Sardana
11	HIPPIE	Production	hippie-scifly-daq.daq.maxiv.lu.se	Scienta/Prodigy
12	MicroMAX	Production	Not started b-v-micromax-ec-2	MxCube Sardana
13	SPECIES	Production	species-scifly-daq.daq.maxiv.lu.se b-v-species-ec-0	Prodigy Sardana
14	BLOCH	Production	bloch-scifly-daq.daq.maxiv.lu.se	Prodigy
15	FemtoMAX	Production	b-v-femtomax-ec-2	Sardana
16	SoftiMAX	Production	softimax-scifly-daq.daq.maxiv.lu.se	Sophie endstation

MAXIV

QUESTIONS?

Simplified MAX IV User Access

