

# EFFICIENT AND AUTOMATED METADATA RECORDING AND VIEWING FOR SCIENTIFIC EXPERIMENTS AT MAX IV

Authors: Daphne van Dijken  
Marco Leorato

Speaker: Vanessa Da Silva

# MAX IV Data Management Principles

Conditions for how scientific data are managed are described in the **MAX IV Experimental Data Policy**, revision 3.

<https://www.maxiv.lu.se/user-access/practical-information/user-policies>

- **Proprietary research** is regulated by separate agreements that prevail over the MAX IV Experimental Data Policy.
- **MAX IV does not guarantee that data is not unintentionally** lost or accessed, nor that it will be available at all times.
- **MAX IV claim no ownership** to Experimental data (raw data and associated metadata), i.e. MAX IV acts as the custodian of experimental data.
- **Raw data is organized and curated** in well-defined formats.
- **Metadata automatically** captured by instruments is either curated within the raw data files or within a associated online catalogue. MAX IV provides its best effort to capture metadata items not automatically captured by an instrument.
- 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**.

# FAIR data policy

F  
Findable



A  
Accessible



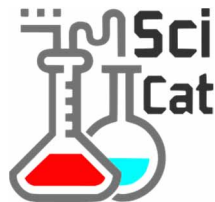
I  
Interoperable



R  
Reusable

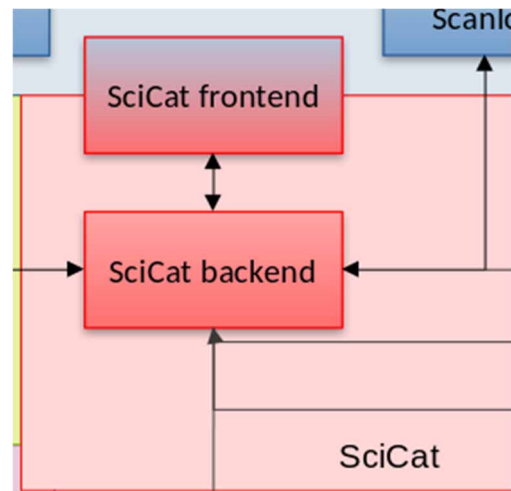


# What is SciCat?



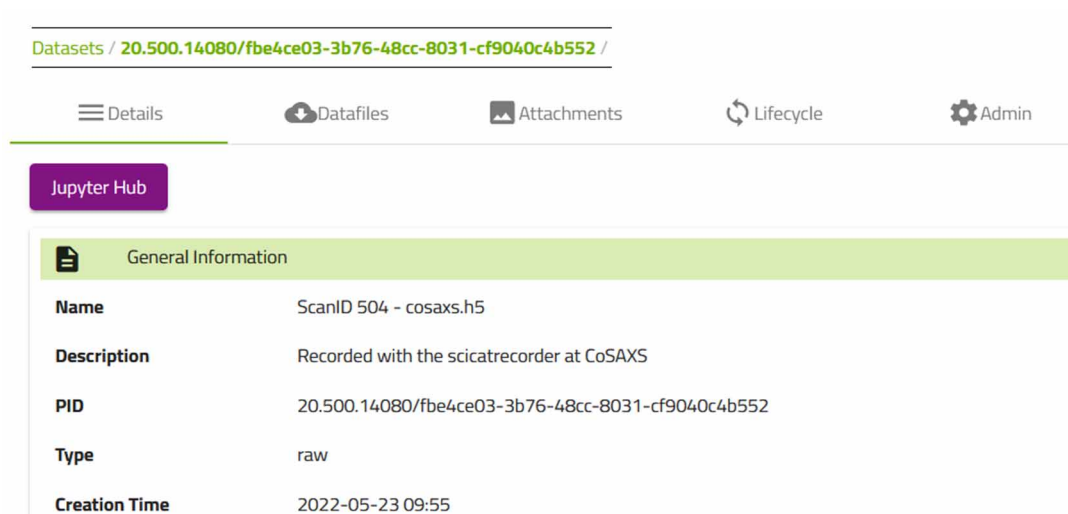
- Open source high performance web based metadata catalog
- Frontend: Angular
- Backend: Nestjs
- Database: MongoDB
- Stores raw and derived datasets, information about proposals, samples, instruments and more

Poster  
THMBCMO02



# Why SciCat?

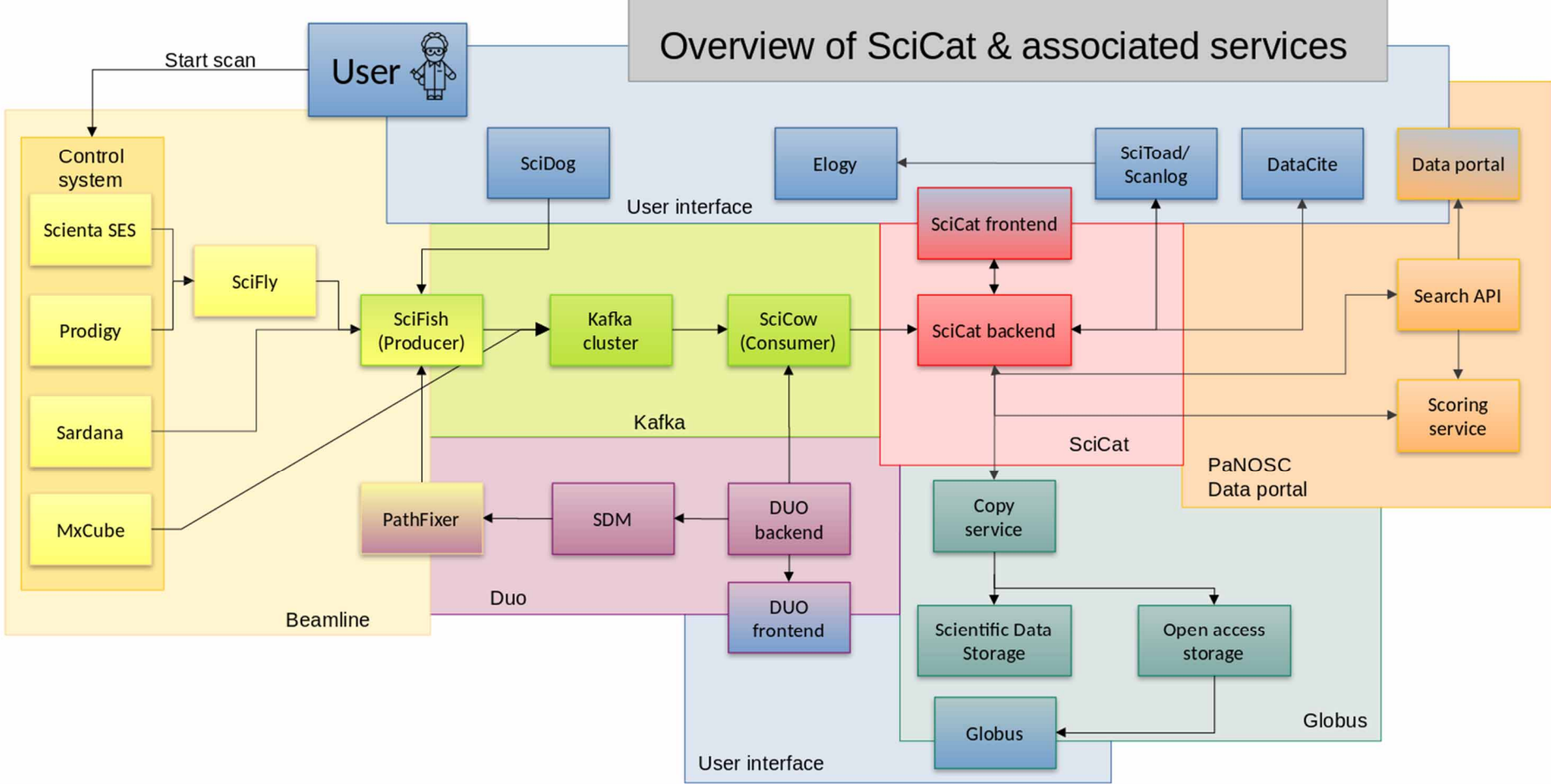
- Collaboration project from PaNOSC
- Designed to support the life-cycle of (meta)data
- More and more data, harder to manage



The screenshot displays the SciCat web interface for a specific dataset. At the top, the breadcrumb path is "Datasets / 20.500.14080/fbe4ce03-3b76-48cc-8031-cf9040c4b552 /". Below this is a navigation bar with five items: "Details" (selected), "Datafiles", "Attachments", "Lifecycle", and "Admin". A purple "Jupyter Hub" button is located below the navigation bar. The main content area is titled "General Information" and contains a table with the following data:

<b>Name</b>	ScanID 504 - cosaxs.h5
<b>Description</b>	Recorded with the scicatrecorder at CoSAXS
<b>PID</b>	20.500.14080/fbe4ce03-3b76-48cc-8031-cf9040c4b552
<b>Type</b>	raw
<b>Creation Time</b>	2022-05-23 09:55

# Overview of SciCat & associated services



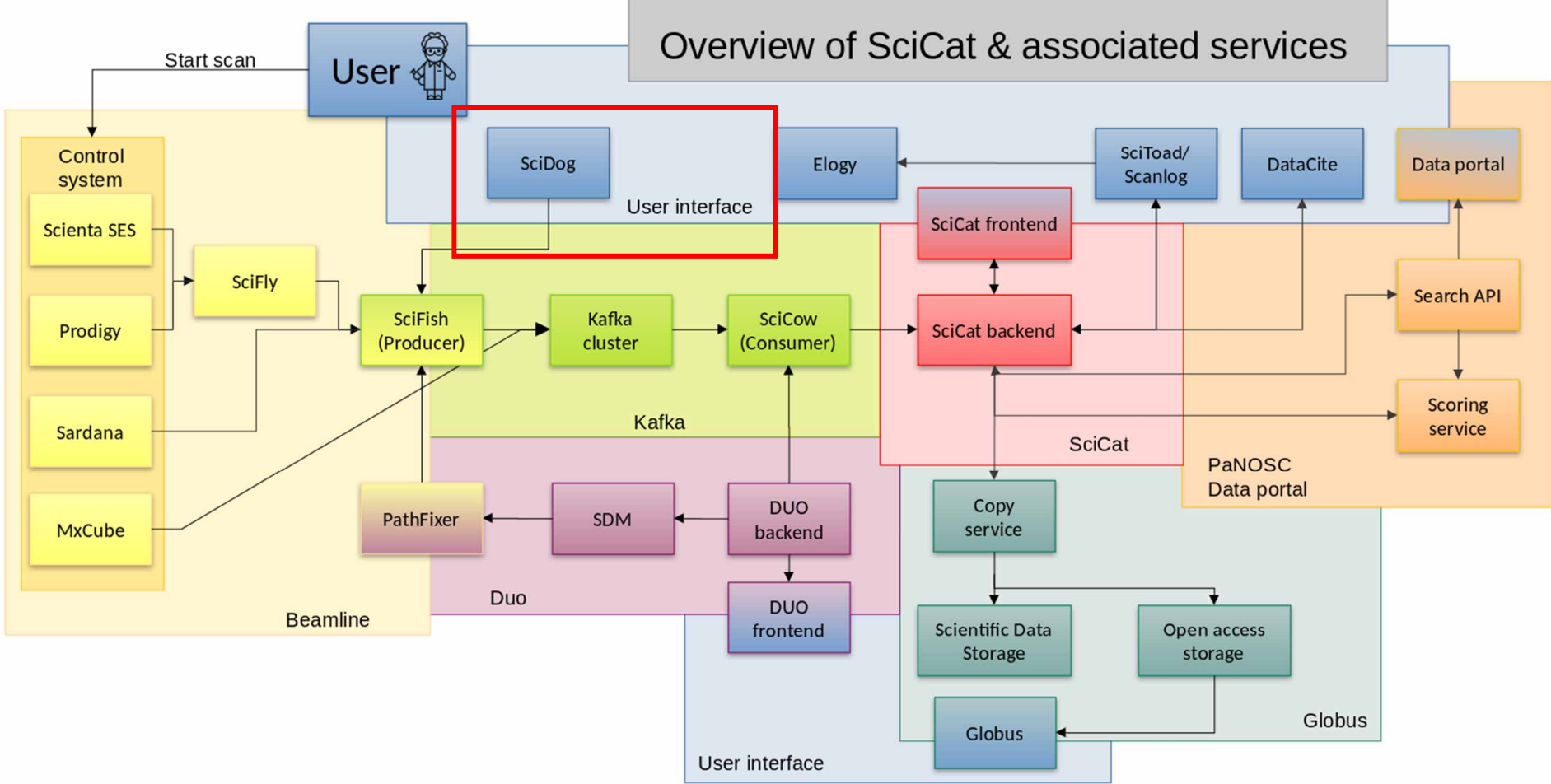
# SciZoo



- SciCat – Scientific Catalog
- SciDog – Scientific (meta)Data organizer GUI
- SciFish – Scientific Form Implementation Service Helper
- SciToad (Scanlog) – Scientific Tool of Attesting (meta)Data
- SciFly – Scientific File Lystener
- SciCow – Scientific Consumer of Wisdom



# Overview of SciCat & associated services





# When a user comes to MAX IV



Scientific (meta)Data Organizer



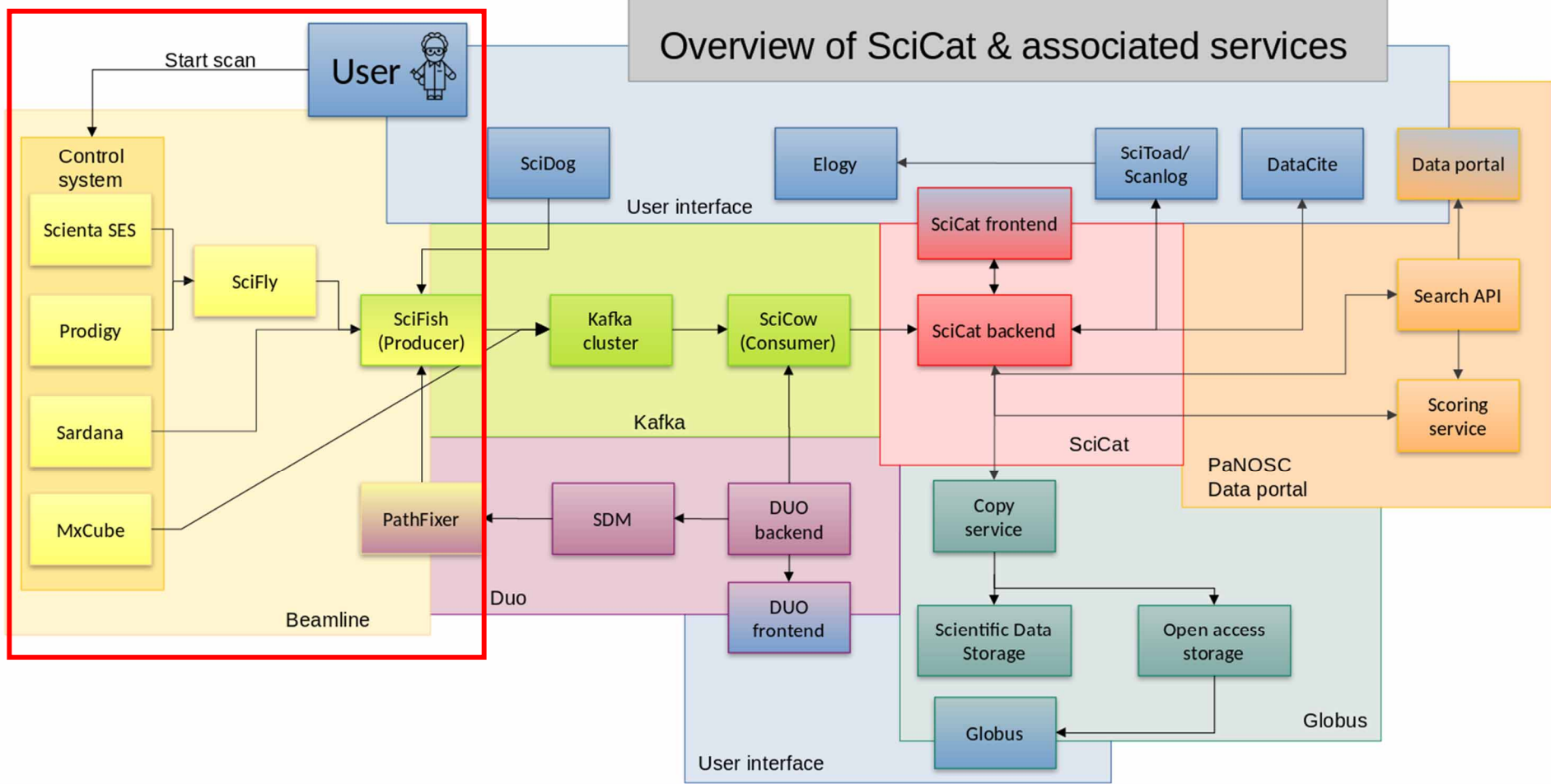
## Scientific Metadata Configuration

Show configuration:

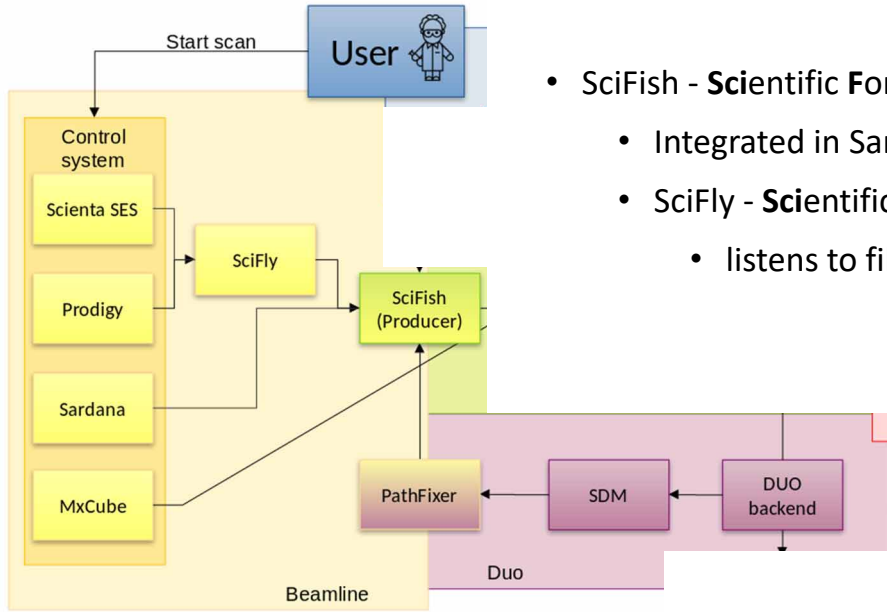
Group	Name	Device	Attribute	Unit
	T02	B310A-E01/DIA/TCO-02	temperature	deg C
	description	b310a-e/ctl/sample	descr	
	well	b310a-e/ctl/sample	well	
	concentration	b310a-e/ctl/sample	concentration	
	T01	B310A-E01/DIA/TCO-01	temperature	degC
	Number images	expchan/eiger_tr_ctrl/1	nbimages	
	Number triggers	expchan/eiger_tr_ctrl/1	nbtriggers	
	WaterBath	b310a-e01/wat/chill-03	temperature	degC

- Use SciDog to configure which metadata should be recorded
- Web-based GUI

# Overview of SciCat & associated services



# When a user runs their acquisitions



- SciFish - **Scientific Form Implementation Service Helper**

- Integrated in Sardana

- SciFly - **Scientific File Lystener**

- listens to files written by Scienta SES or Prodigy

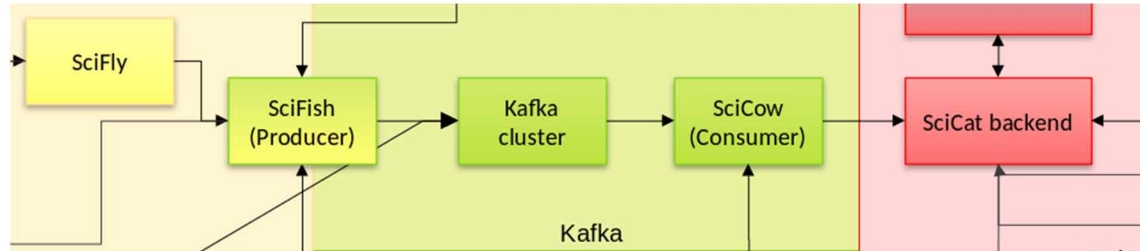
- Metadata is automatically harvested from various places

- From DUO (Digital User Office) for proposal information
- From Tango Devices for information about specific parameters
- From Sardana, the orchestration system, for information about a scan

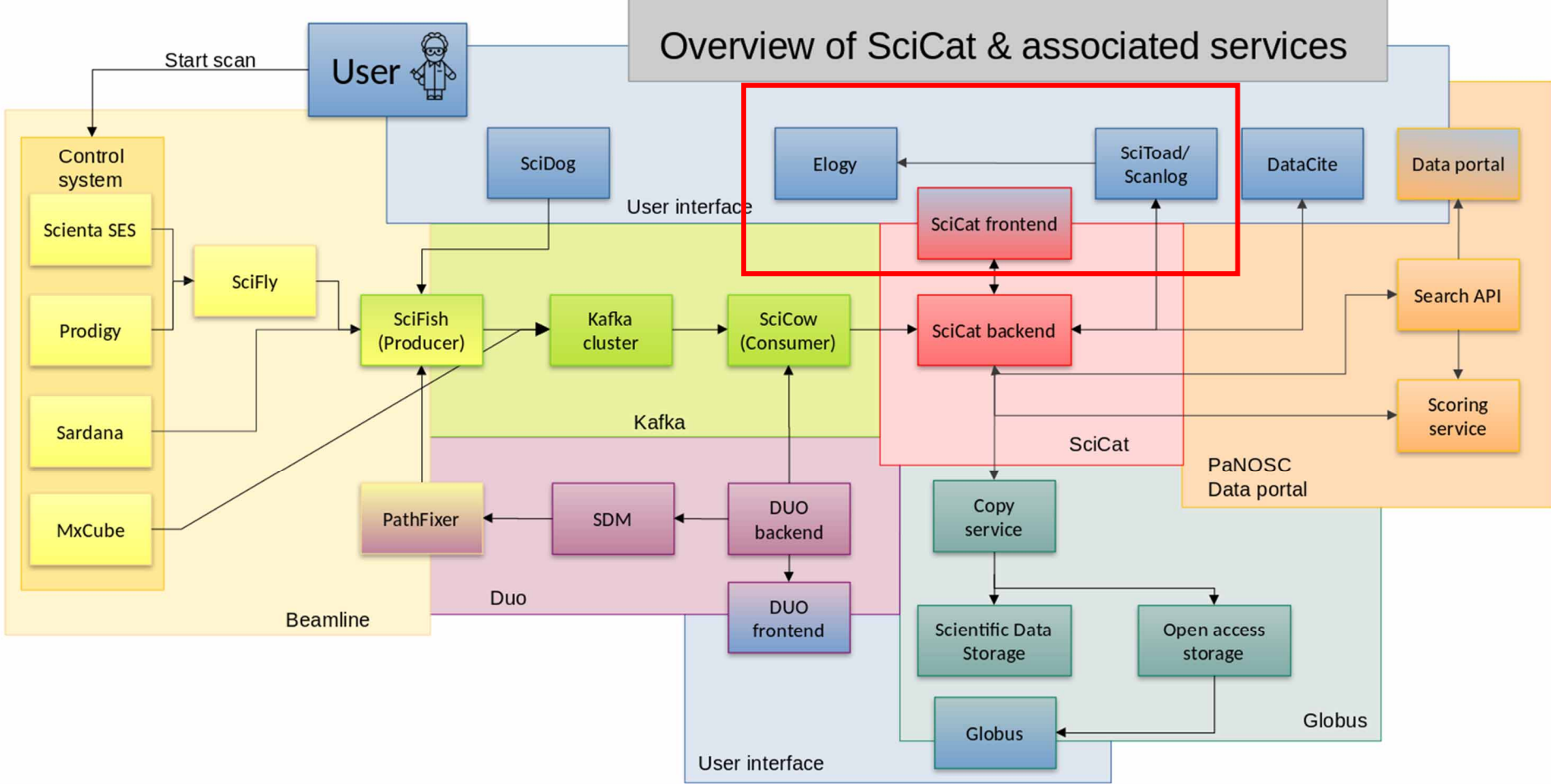
# When a user runs their acquisitions

SciCoW - **Scientific Consumer of Wisdom**

This metadata is automatically gathered and put in the database by a Kafka Consumer

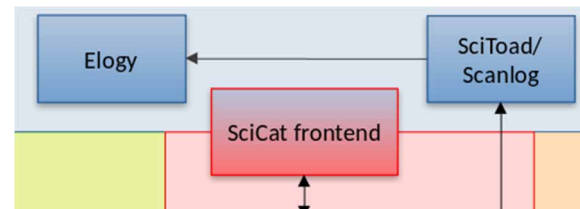


# Overview of SciCat & associated services



# When a user runs their acquisitions

- The users can create a customized table in Scanlog/SciToad, where they can view their metadata and optionally add comments, upload images and rate their dataset



**MAXIV SciCat Scanlog - CoSAXS** Logged in as ingestor [Log out](#)

Showing scans from 2021-04-01 00:00:00 till 2021-04-14 00:00:00 20190679

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

ScanID	Start time	PI	Name	Scan	Detectors	Data path	SciCat	Comment	Upload Image	Images	Data Quality Metrics
<input type="checkbox"/> 64	2021-04-09 10:55:17.253	car146@gmail.com	test12.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 62	2021-04-08 18:04:58.909	car146@gmail.com	test12.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>	outstanding, admirable	<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★★★★	✘
<input type="checkbox"/> 61	2021-04-08 18:04:54.440	car146@gmail.com	test11.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★★★★	✘
<input type="checkbox"/> 60	2021-04-08 18:04:36.019	car146@gmail.com	test10.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 58	2021-04-08 17:58:21.012	car146@gmail.com	data/test	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 57	2021-04-08 17:57:55.882	car146@gmail.com	data/test	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>	execrable, appalling	<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★★★★	✘
<input type="checkbox"/> 53	2021-04-08 17:53:46.643	car146@gmail.com	test7.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 49	2021-04-08 17:50:14.197	car146@gmail.com	test6.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 47	2021-04-08 17:48:42.772	car146@gmail.com	test5.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	
<input type="checkbox"/> 41	2021-04-08 17:30:59.193	car146@gmail.com	test3.h5	ascan mot01 0.0 1.0 1.0 0.0	ct01; ct02; ct03; ct04	/data/test	<a href="#">Link</a>		<a href="#">Browse...</a> No files selected.	<a href="#">Show</a> ★★	

Rows per page: 10 1-10 of 19 < > >>

Title:

Logbook:

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

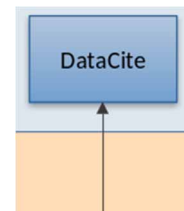
No data selected

[Export to Elogy](#)

[Export to CSV](#)

# When the user wants to publish their data

Datasets are given a unique identifier, a DOI, to be traceable and referenceable in papers



Reference Lysozyme data from the first kilohertz serial crystallography experiments with the JUNGFRÄU 4M detector at a 4th generation synchrotron source

Filip Leonarski, Jie Nan, Zdenek Matej, Quentin Bertrand, Antonia Furrer, Ishkhan Gorgisyan, Monika Bjelčić, Michal Kepa, Hannah Glover, Viktoria Hinger, Thomas Eriksson, Aleksander Cehovin, Mikel Eguirau, Piero Gasparotto, Aldo Mozzanica, Tobias Weinert, Ana Gonzalez, Jörg Standfuss, Meitlan Wang, Thomas Ursby, Florian Dworkowski, MAX IV Laboratory, Lund University (2023)

## Abstract

The experiment is a collaboration between MAX IV and PSI to prove the feasibility to use the high frame-rate state of the art integrating pixel detector (JUNGFRÄU) for serial crystallography experiments at a 4th generation synchrotron light source. A PSI-provided JUNGFRÄU 4M prototype detector was installed at the BioMAX beamline. The serial synchrotron crystallography (SSX) experiment was using the MAX IV high-viscosity extruder (HVE) for sample delivery. Two reference datasets were collected at 1 kHz and 2 kHz frame rates. The third reference dataset was collected with frame rate 300 kHz using 1 kHz detector settings and adjusted flux and jet parameters in order to preserve constant radiation dose.

## Publication details

DOI <https://doi.org/10.48391/boc36bb8-a00c-4519-8d6c-08d5ca60a313>

Resource Type raw

Size 1 TB

Related Publications F. Leonarski et al. (2023) - being submitted

## Datasets

Data Description Reference Lysozyme data collected within the first kilohertz serial crystallography experiments with the JUNGFRÄU 4M detector at a 4th generation synchrotron source. The collection contains in total 3 datasets, acquired at 100 Hz, 1 kHz and 2 kHz respectively.

[20.500.14080/773e5762-120e-4f3b-87b8-f89e72d32ee](https://doi.org/10.48391/boc36bb8-a00c-4519-8d6c-08d5ca60a313)

[20.500.14080/18902a4-d4c2-4c4a-8e99-92c27adef976](https://doi.org/10.48391/boc36bb8-a00c-4519-8d6c-08d5ca60a313)

[20.500.14080/7696a219-36f8-4bec-a188-95849aeac05](https://doi.org/10.48391/boc36bb8-a00c-4519-8d6c-08d5ca60a313)

## Public Data Repository Dashboard

Items per page: 20

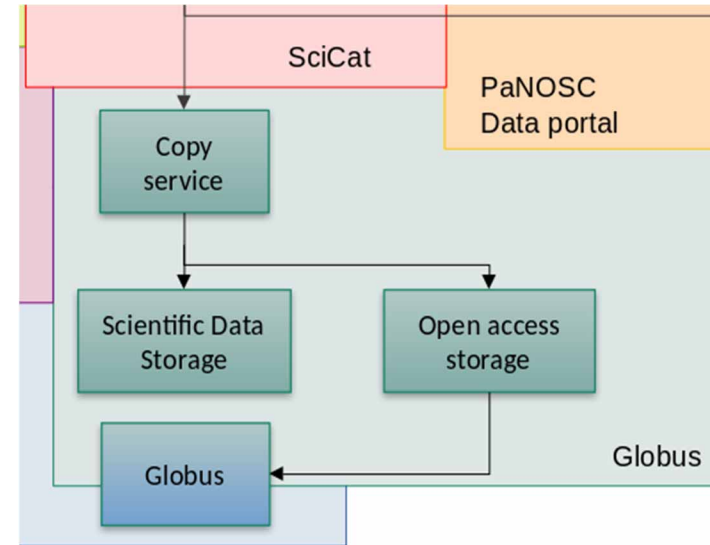
1 - 3 of 3

< >

<input type="checkbox"/>	Title	Registered Time	Creator	Doi
<input type="checkbox"/>	Remote experiments at Bio-MAX	2022-05-18 09:25	Ana Gonzalez	10.48391/91a3a886-df6f-434a-9a1d-7304a0061e00
<input type="checkbox"/>	OSC collection in Biomax for Sample 47910	2022-05-18 10:54	Ana Gonzalez	10.48391/b0179486-47c5-4e5c-8301-cde7bfc2ce1
<input type="checkbox"/>	Reference Lysozyme data from the first kilohertz serial crystallography experiments with the JUNGFRÄU 4M detector at a 4th generation synchrotron source	2023-04-25 17:54	Filip Leonarski, Jie Nan, Zdenek Matej, Quentin Bertrand, Antonia Furrer, Ishkhan Gorgisyan, Monika Bjelčić, Michal Kepa, Hannah Glover, Viktoria Hinger, Thomas Eriksson, Aleksander Cehovin, Mikel Eguirau, Piero Gasparotto, Aldo Mozzanica, Tobias Weinert, Ana Gonzalez, Jörg Standfuss, Meitlan Wang, Thomas Ursby, Florian Dworkowski	10.48391/boc36bb8-a00c-4519-8d6c-08d5ca60a313

# Retrieving data

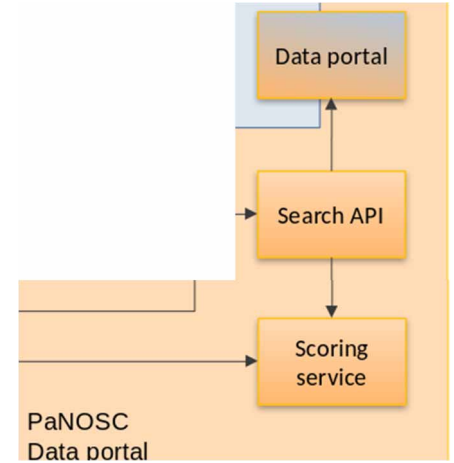
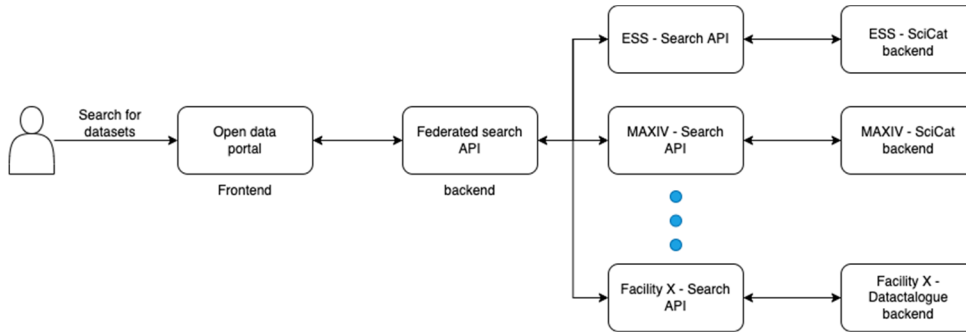
- A download link can be retrieved through SciCat
- Datafiles can be transferred in Globus





# Integration with PaNOSC Data Portal

- Loopback search API that is connected to SciCat backend
- Can see published scans from multiple facilities and filter based on parameters defined in the (scientific) metadata
- Scoring service, parameters that calculate how relevant the item is compared to the search term given by the user
- Published scans are available for anyone



# Future

- Implement data processing and analysis pipelines (JupyterHub)
- Add more information in SciCat such as sample and instrument
- Publishing of datasets with DOI in more data portals (OpenAIRE, B2Find)

# Thank you!

**Members of MAX IV Software Group**  
Daphne van Dijken [daphne.van\\_dijken@maxiv.lu.se](mailto:daphne.van_dijken@maxiv.lu.se)  
Marco Leorato [marco.leorato@maxiv.lu.se](mailto:marco.leorato@maxiv.lu.se)  
Vanessa Silva [vanessa.silva@maxiv.lu.se](mailto:vanessa.silva@maxiv.lu.se)

