

# SciCat Development



# SciCat

Max Novelli

[max.novelli@ess.eu](mailto:max.novelli@ess.eu)

SciCat Project Leader

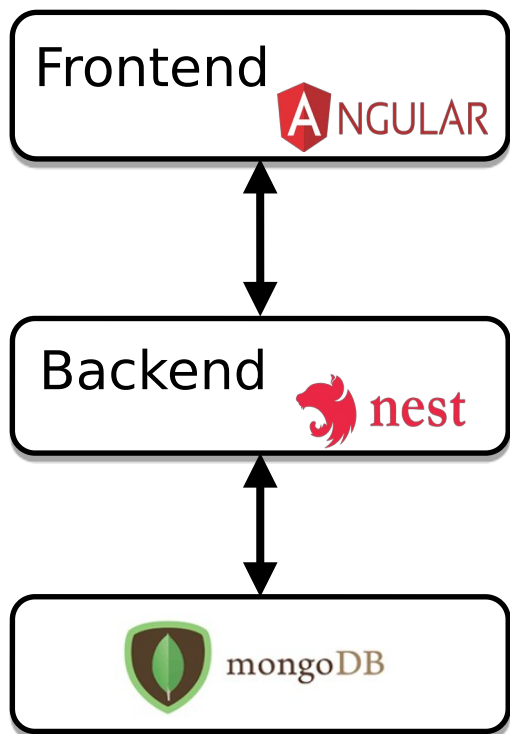
Senior Data Curation Scientist  
European Spallation Source, DMSC  
Copenhagen, Denmark

MXCuBE ISPyB 2024, Trieste, Italy, 2024/11/20



# SciCat Project

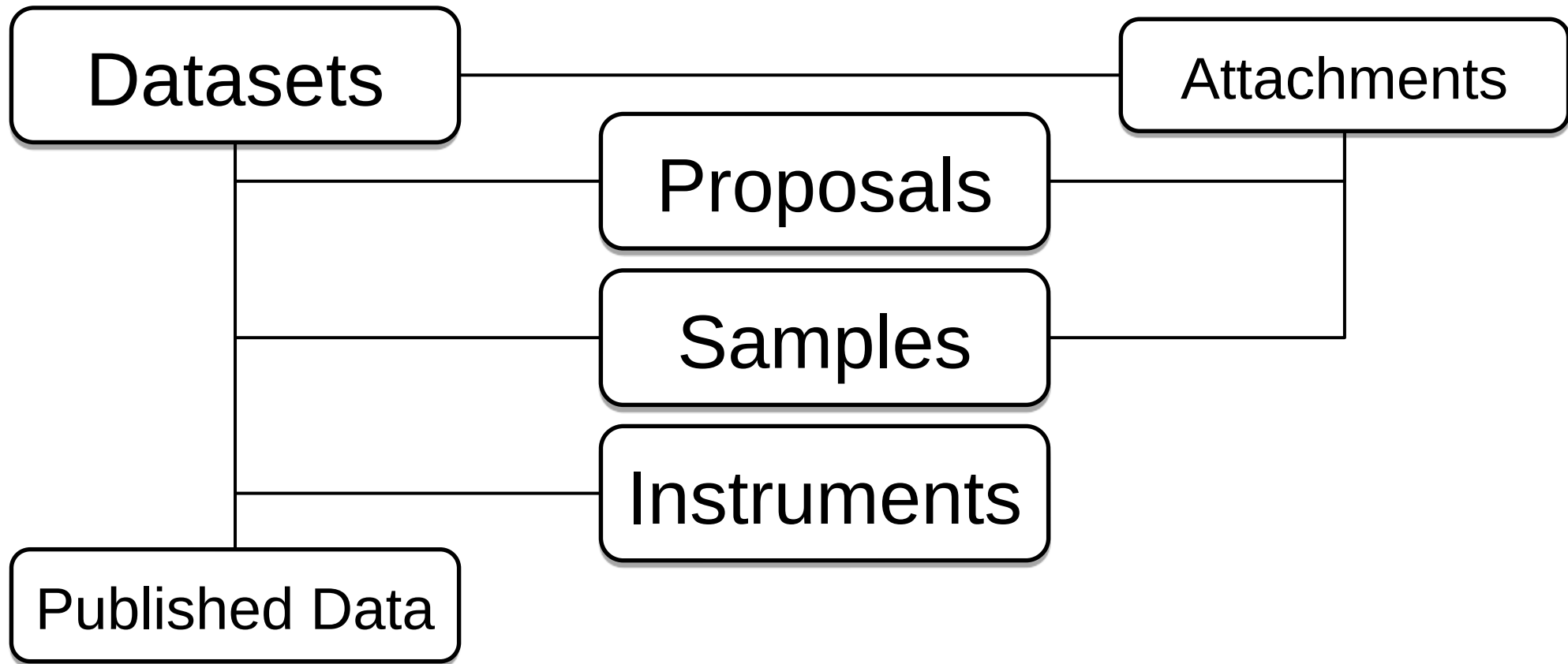
## Open Source Scientific Metadata Catalog



<https://scicatproject.github.io/>  
<https://github.com/scicatproject>  
<https://scicatproject.github.io/documentation>

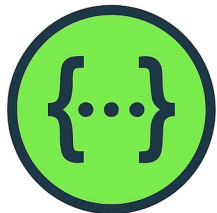
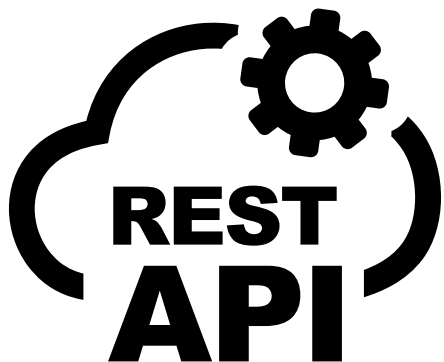


# SciCat Sub-Systems





# Backend Interface



Swagger™

## SciCat backend API undefined OAS 3.0

This is the API for the SciCat Backend

Authorize

auth

users

user identities

datasets

**POST** /api/v3/datasets It creates a new dataset which can be a raw or derived one.

**GET** /api/v3/datasets It returns a list of datasets.

**POST** /api/v3/datasets/isValid It validates the dataset provided as input.

**GET** /api/v3/datasets/fullquery It returns a list of datasets matching the query provided.

**GET** /api/v3/datasets/fullfacet It returns a list of dataset facets matching the filter provided.

**GET** /api/v3/datasets/metadataKeys It returns a list of metadata keys contained in the datasets matching the filter provided.

**GET** /api/v3/datasets/findOne It returns the first dataset found.

**GET** /api/v3/datasets/count It returns the number of datasets.

**GET** /api/v3/datasets/{pid} It returns the dataset requested.

**PATCH** /api/v3/datasets/{pid} It partially updates the dataset.

**PUT** /api/v3/datasets/{pid} It updates the dataset.

**DELETE** /api/v3/datasets/{pid} It deletes the dataset.





# Frontend Interfaces

SciCat STAGING

Help About Sign in

Datasets /

Search

Items per page: 25 1 - 25 of 2896

**Filters and Conditions**

- Location
- Pid
- Group
- Type
- Keyword
- Start Date - End Date
- Text
- 

Pid	Name	Run No.	Source Folder	Size	Start Time	Type	Image	Proposal ID
20.500.12269/a76901e4-7e3a-4bd1be09-6845f1f29b83	Dataset 20.500.12269/a76901e4-7e3a-4bd1-be09-6845f1f29b83 for proposal 964281 run 11303		...597001/raw	0 B	2024-07-22 Mon 09:07	raw		964281
20.500.12269/8ea985c555-4b0a-9070-56c	This is test		.../nfs/test	0 B	2023-10-05 Mon 14:18	derived		
20.500.12269/7d4572ab2d-bdc9f440cb30	Fe304		...ction/test	14 KiB	2020-02-20 Thu 11:30	raw		
20.500.12269/2d5af6	V20 sample data		.../nfs/disk1	0 B	2020-02-06 Thu 14:02	derived		
20.500.12269/2511ni	Last Neutrons Ever at HZB.	2511	...v20/YC7S25	203 MiB	2019-12-11 Wed 13:48	raw		YC7S25
20.500.12269/2510ni	WFM Low Res messing with chopper 1 and 2, slit2 pinhole, slit3 fully open.	2510	...v20/YC7S25	81 MiB	2019-12-11 Wed 13:37	raw		YC7S25
20.500.12269/2509ni	slitscan, slit2 scanning, slit3 fully open	2509	...v20/YC7S25	116 MiB	2019-12-11 Wed 12:54	raw		YC7S25
20.500.12269/2508ni	slitscan, slit2 scanning, slit3 fully open	2508	...v20/YC7S25	38 MiB	2019-12-11 Wed 12:49	raw		YC7S25
20.500.12269/2507ni	slitscan, slit2 scanning, slit3 fully open	2507	...v20/YC7S25	2 KiB	2019-12-11 Wed 12:48	raw		YC7S25

Datasets list



# Frontend Interfaces

## Datasets details

SciCat **ESS STAGING** [Help](#)

Datasets / 20.500.12269/268e616e-a5bc-11ef-a3ef-fa163e9aae0a /

[Details](#) [Datafiles](#) [Related Datasets](#) [Logbook](#) [Attachments](#)

### General Information

**Name** loki

**Description** Dataset: 20.500.12269/268e616e-a5bc-11ef-a3ef-fa163e9aae0a. Proposal: 443503. Sample: unknown. Instrument: LoKI on CODA. Data file writer job:268e616e-a5bc-11ef-a3ef-fa163e9aae0a

**PID** 20.500.12269/268e616e-a5bc-11ef-a3ef-fa163e9aae0a [Copy](#)

**Type** raw

**Creation Time** 2024-11-18 15:57

**Keywords** [CODA](#) [LOKI](#) [CODA / LOKI](#) [POST\\_CURATED](#)

### Creator Information

**Owner** Massimiliano Novelli

**Principal Investigator** Massimiliano Novelli

**Contact Email** [max.novelli@ess.eu](mailto:max.novelli@ess.eu)

**Owner Group** 443503

**Access Groups** ecdc,sims,dram,ess,scientific information management systems group,ess employees

### File Information

**Source Folder** /ess/data/coda/2024/443503

### Related Documents

**Proposal** [CODA Proposal #3](#)

**Instrument** [coda](#)

**Creation Location** ESS/CODA/LOKI

### Scientific Metadata

[View](#)

Name	Value
Run Number	33909
Start Time	2024-11-18T14:48:19.109Z
End Time	2024-11-18T14:57:33.342Z
Job Id	268e616e-a5bc-11ef-a3ef-fa163e9aae0a
Acquisition Team Members	Anders Pettersson, Fredrik Bolmsten, George O'Neill, Jonas Petersson, Massimiliano Novelli, Matt Clarke
Simulated Instrument Name	LoKI
Simulated Instrument Id	20.500.12269/58756d47-932f-4b73-8544-7dccb0f9827e
Simulated Instrument Url	<a href="https://staging.scicat.ess.eu/instruments/20.500.12269%2F58756d47-932f-4b73-8544-7dccb0f9827e">https://staging.scicat.ess.eu/instruments/20.500.12269%2F58756d47-932f-4b73-8544-7dccb0f9827e</a>



# Frontend Interfaces

## Proposal

[Proposals](#) / [443503](#) /

[Details](#) [Datasets](#) [Logbook](#)

### General Information

<b>Title</b>	CODA Proposal #3
<b>Abstract</b>	A proposal to continue running our test on CODA
<b>Identifier</b>	443503

### Creator Information

<b>Main proposer</b>	<a href="#">Massimiliano Novelli</a>
<b>Principal investigator</b>	<a href="#">Massimiliano Novelli</a>



# Frontend Interfaces

# Samples

Samples / [03792382-84d6-40ec-b909-e0959752b0a0](#) /

[Details](#)

[Attachments](#)

## General Information

<b>Description</b>	Ni_Ti multilayer
<b>Sample ID</b>	03792382-84d6-40ec-b909-e0959752b0a0
<b>Creation Time</b>	2024-10-08 10:37

## Creator Information

<b>Sample Owner</b>	Proposal 533929
<b>Owner Group</b>	533929
<b>Access Groups</b>	ecdc,sims,dram,ess,scientific information management systems group,ess employees

## Characteristics

[View](#)

[Edit](#)

Name	Value	Unit
Sample Formula	Ni_Ti	--
Sample State	Solid	--





# Development



Jobs



ES Integration

Search UI



SciCatLive

SDK



Proposals

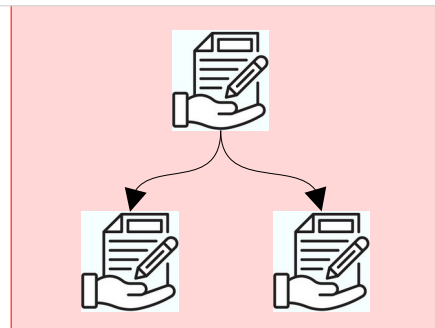
Proposals

Flat



Hierarchical

Dedicated Metadata



Proposals / 443503 /

[Details](#) [Datasets](#) [Logbook](#)



**General Information**

**Title** CODA Proposal #3  
**Abstract** A proposal to continue running our test on CODA  
**Identifier** 443503

**Creator Information**

**Main proposer** Massimiliano Novelli  
**Principal investigator** Massimiliano Novelli

**Parent Proposal**



**Characteristics**

[View](#) [Edit](#)

Name	Value	Unit
Sample Formula	Ni_Ti	--
Sample State	Solid	--



# Future Development

## Dataset Details View

- Flexible
- Customizable
- Selectable

The screenshot displays the 'Dataset Details View' for a specific dataset. The interface is organized into several sections:

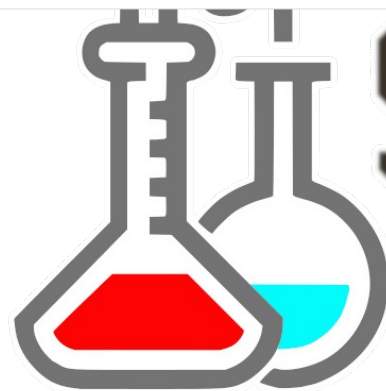
- General Information:** Shows the dataset ID (20.500.12269/268e616e-a5bc-11ef-a3ef-f4163e9aa0a), creation time (2024-11-18 15:57), and keywords (CODA, LDKI, CODA / LDKI, POST\_CURATED).
- Creator Information:** Lists the owner (Massimiliano Novelli), principal investigator (Massimiliano Novelli), contact email (max.novelli@ess.eu), owner group (443903), and access groups (erc\_sims,dram,ess,scientific information management systems group,ess employees).
- File Information:** Shows the source folder (/ess/ista/coda/2024/443903).
- Related Documents:** Lists the proposal (CODA Proposal #3), instrument (coda), and creation location (ESS/CODA/LDKI).
- Scientific Metadata:** A table with columns for Name, Value, and Unit, containing details such as Run Number (33909), Start Time (2024-11-18T14:48:19.109Z), End Time (2024-11-18T14:57:33.342Z), Job ID (268e616e-a5bc-11ef-a3ef-f4163e9aa0a), Acquisition Team Members (Anders Pettersson, Fredrik Bolmsten, George O'Neill, Jonas Pettersson, Massimiliano Novelli, Matt Clarke), Simulated Instrument Name (Loki), Simulated Instrument ID (20.500.12269/5875647-932f-4b73-8544-76d80f9827e), and Simulated Instrument URI (https://staging.socat.ess.eu/instruments/20.500.12269/5875647-932f-4b73-8544-76d80f9827e).



# Integration



ISPyB - SciCat



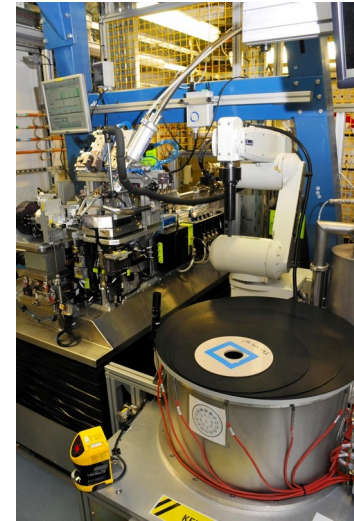
SciCat



# Sample Management



- Sample Shipment ✓
- Sample Declaration ✓
- Puck Layout ✓
- Metadata Catalogue ✓





# ISPyB Sample

101001001010

## Shipment

Information [Transport history](#)

Name: **LP3\_210702** Date:  [Send shipment to the facility](#) Comment:  [Edit](#)

Beamline:  Status: **processing** [Delete](#)

From: **Knecht-Lund University** Courier company: **-1** Allowed Reimb. parcels: **0**

Return address: **Knecht-Lund University** Billing Reference:  Fedex Reference:

# Shipment Management

Content (1 Parcels - 85 Samples - 77 Measured)

[Add Parcel](#) [Import from CSV](#) [Export PDF View](#)

#1

Name: **LP3\_dewar1**

Status: **processing**

Location:

Storage: **Dry shipper**

Comments: User will collect dewar after beamtime.

## Prepare Experiment

1 Select Shipment

2 Load Sample Changer

### Loaded or to be Loaded on MxCube

Shipment	Parcel	Container	Container type	Beamline	Cell	Position
LP3_210702	LP3_dewar1	MAX0143 (16 samples)	UNIPUCK	BioMAX	1	1
LP3_210702	LP3_dewar1	MAX0144 (16 samples)	UNIPUCK	BioMAX	1	2
LP3_210702	LP3_dewar1	MAX0145 (14 samples)	UNIPUCK	BioMAX	1	3
LP3_210702	LP3_dewar1	MAX0146 (12 samples)	UNIPUCK	BioMAX	1	4

### BioMAX (ISARA)

Unload SC

Container: MAX0146  
SC Location: 4

Unload all

Unload #4

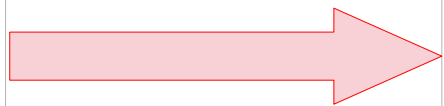
Protein acronym: Nsp10s  
Sample: NSP10S\_562\_A1a

# Puck Layout



# Sample: ISPyB to SciCat <sup>(1)</sup>

Sample



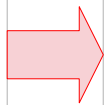
Sample Item

Flat



Hierarchical

Single Type



Multiple Types



# Sample: ISPyB to SciCat (2)



## Shipment

- All Sample Items
- Different Types
- Properties in Metadata
- Use metadata schema



## Sample



## PUCK



## Sample Charger





# ISPyB Datasets

## Dataset Overview


14 Data Collections | Energy Scans | 1 Fluorescence Spectra

PDF summary | RTF summary | CSV summary | PDF analysis | RTF analysis | enter search prefix, sample, protein or filePath

OSC 11-05-2020 20:00:35  
/data/visitors/biomax/20200556/20200511/raw/Nsp10/Nsp10-285\_G12-3I/

Summary | Beamline Parameters | Data Collections 1 | Sample | Last Collect Results 5 | Workflow

Workflow	Res. (corner)	2.00 Å ( Å )
Protein	Nsp10	En. (Wave.) 12.700 keV (0.9763 Å)
Sample	285_G12-3I	Phi range 0.10 °
Prefix	Nsp10-285_G12-3I	Phi start (total) 90.00 ° (180°)
Run #	1	Exposure Time 0.011 s
# Images (Total)	1800 (1800)	Flux start 1.24e+12 ph/sec
Transmission	40.3 %	Flux end ph/sec



Comments:

14 Data Collections | Energy Scans | 1 Fluorescence Spectra

PDF summary | RTF summary | CSV summary | PDF analysis | RTF analysis | enter search prefix, sample, protein or filePath

OSC 11-05-2020 20:00:35  
/data/visitors/biomax/20200556/20200511/raw/Nsp10/Nsp10-285\_G12-3I/

Summary | Beamline Parameters | Data Collections 1 | Sample | Last Collect Results 5 | Workflow

Workflow	Res. (corner)	2.00 Å ( Å )	Beamline name	Synchrotron name	MAXIV	Focusing optics	KB mirrors	Detector Type	Hybrid pixel direct co	
Protein	Nsp10	En. (Wave.) 12.700 keV (0.9763 Å)	Detector Distance	286.00 mm	Synchrotron filling mode	Variable TopUp/Decay	monochromator type	Si(111)	Detector Model	EIGER
Sample	285_G12-3I	Phi range 0.10 °	X Beam	2100.77 mm	Synchrotron Current	249.8 mA	Beam shape	ellipse	Manufacturer	DECTRIS
Prefix	Nsp10-285_G12-3I	Phi start (total) 90.00 ° (180°)	Y Beam	2121.33 mm	Undulator types		Beam size at Sample Hor (Vert)	50 (50) µm	Pixel Size Hor (Vert)	75 (75) µm
Run #	1	Exposure Time 0.011 s	Kappa	-9999.00	Undulator gaps	5.01 mm	Beam divergence Hor (Vert)	6 (104) µrad	Polarisation	0.99
# Images (Total)	1800 (1800)	Flux start 1.24e+12 ph/sec	Phi	-9999.00						
Transmission	40.3 %	Flux end ph/sec								

Comments:

## Dataset Metadata



# Datasets: ISPyB to SciCat

14 Data Collections | Energy Scans | 1 Fluorescence Spectra


PDF summary | RTF summary | CSV summary | PDF analysis | RTF analysis | enter search prefix, sample, protein or filePath

OSC 11-05-2020 20:00:35  
/data/visitors/blomax/20200556/20200511/raw/Nsp10/Nsp10-285\_G12-3I/

Summary | Beamline Parameters | Data Collections 1 | Sample | Last Collect Results 6 | Workflow

Workflow	Res. (corner)	2.00 Å (Å)	
Protein	Nsp10	En. (Wave.)	12.700 keV (0.9763 Å)
Sample	285_G12-3I	Phi range	0.10 °
Prefix	Nsp10-285_G12-3I	Phi start (total)	90.00 ° (180°)
Run #	1	Exposure Time	0.011 s
# Images (Total)	1800 (1800)	Flux start	1.24e+12 ph/sec
Transmission	40.3 %	Flux end	ph/sec

Comments:



Related Datasets | Logbook | Attachments | Lifecycle

Public

- raw\_sample\_image
- averaged\_dark\_current\_image
- averaged\_open\_beam\_image
- normalized\_sample\_image
- angle\_1\_summary\_image
- angle\_2\_summary\_image

500.12269/90d87f9a-950f-11ef-9101-fa163e943f63, Reconstruction b22f537d7a6c

500.12269/90d87f9a-950f-11ef-9101-fa163e943f63, Proposal 964281, Run 11465, Reconstruction short f6831142f-d2f7-4c38-97f6-b22f537d7a6c

99f0-4d2042d027ab

YMIR | LIGHT TOMOGRAPHY | LEGO SAMPLE

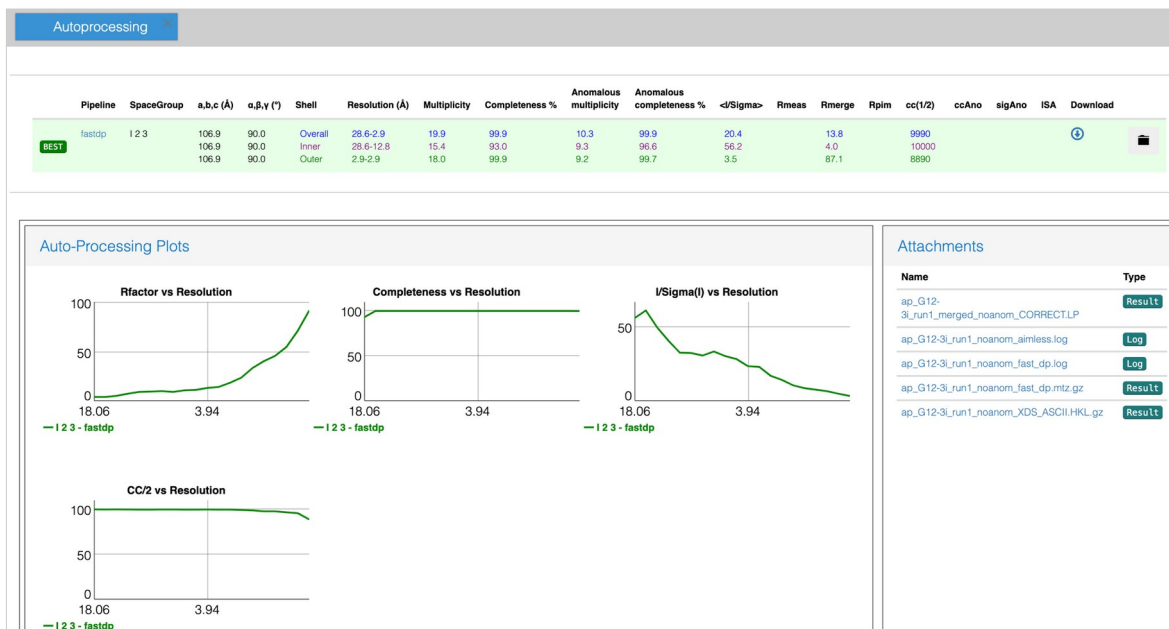
Scientific Metadata

Name	Value	Unit
Raw Proposal Id	964281	--
Raw Run Number	11465	--
Reconstruction Id	6831142f-d2f7-4c38-97f6-b22f537d7a6c	--
Reconstruction Short Id	b22f537d7a6c	--
Image Detector Name	orca	--
Rotation Motion Sensor Name	motion_cabinet_2	--



# ISPyB Autoprocessing

Upon data collection,  
automatic reduction pipelines are triggered

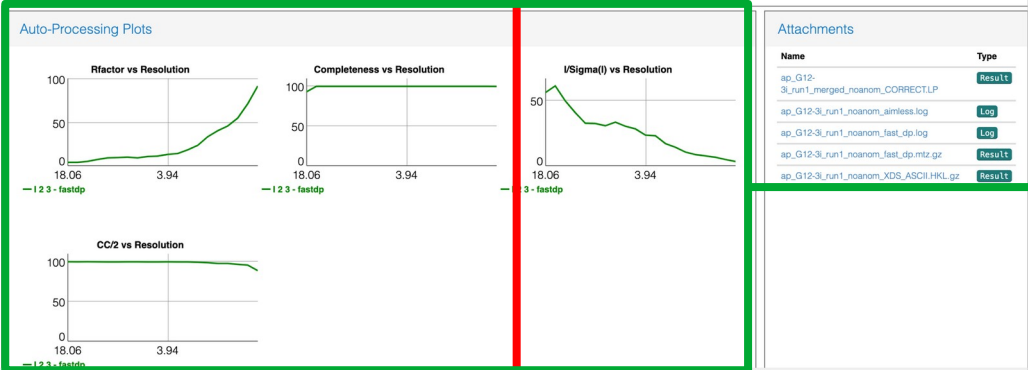




# Autoproc.: ISPyB to SciCat

Autoprocessing

Pipeline	SpaceGroup	a,b,c (Å)	$\alpha,\beta,\gamma$ (°)	Shell	Resolution (Å)	Multiplicity	Completeness %	Anomalous multiplicity	Anomalous completeness %	$\langle I \rangle / \sigma(I)$	Rmeas	Rmerge	Rpim	cc(1/2)	ccAno	sigAno	ISA	Download
BEST	fastdp	1 2 3	106.9 90.0 106.9 90.0 106.9 90.0	Overall Inner Outer	28.6-2.9 28.6-12.8 2.9-2.9	19.9 15.4 18.0	99.9 93.0 99.9	10.3 9.3 9.2	99.9 96.6 99.7	20.4 56.2 3.5	13.8 4.0 87.1	9990 10000 8890						



Details | Datafiles | Related Datasets | Logbook | Attachments | Lifecycle

### General Information

**Name** Reconstructed data for raw dataset 20.500.12269/90d87f9a-950f-11ef-9101-fa163e943f63, Reconstruction b22f537d7a6c

**Description** Reconstructed data for raw dataset 20.500.12269/90d87f9a-950f-11ef-9101-fa163e943f63, Proposal 964281, Run 11465, Reconstruction short id b22f537d7a6c, Reconstruction full id 6831142f-d2f7-4c38-97f6-b22f537d7a6c

**PID** 20.500.12269/78e1c18e-94a1-41cc-99f0-4d2042d027ab

**Type** derived

**Creation Time** 2024-10-28 10:40

**Keywords** TOMOGRAPHY RECONSTRUCTION YMIR LIGHT TOMOGRAPHY LEGO SAMPLE

**Scientific Metadata**

Name	Value	Unit
Raw Proposal Id	964281	--
Raw Run Number	11465	--
Reconstruction Id	6831142f-d2f7-4c38-97f6-b22f537d7a6c	--
Reconstruction Short Id	b22f537d7a6c	--
Image Detector Name	orca	--
Rotation Motion Sensor Name	motion_cabinet_2	--

raw\_sample\_image  
averaged\_dark\_current\_image  
averaged\_open\_beam\_image  
normalized\_sample\_image  
angle\_1\_summary\_image  
angle\_2\_summary\_image







# Roadmap

## Action items

- Data mapping
- SciCat backend features
- Refactor current ISPyB UIs  
to use SciCat BE leveraging SDK
- Dedicated datasets details view in SciCat UI



All effort available within the SciCat project is currently allocated.

SciCat collaborators and ESS will be excited to assist, support and manage the effort.

Additional resources needs to be found and allocated for this integration





# Thank you

Massimiliano (Max) Novelli  
*max.novelli@ess.eu*



EUROPEAN  
SPALLATION  
SOURCE

## Thanks to all SciCat contributors