

## Uniform Data Model For SciCat At DESY FS-EC

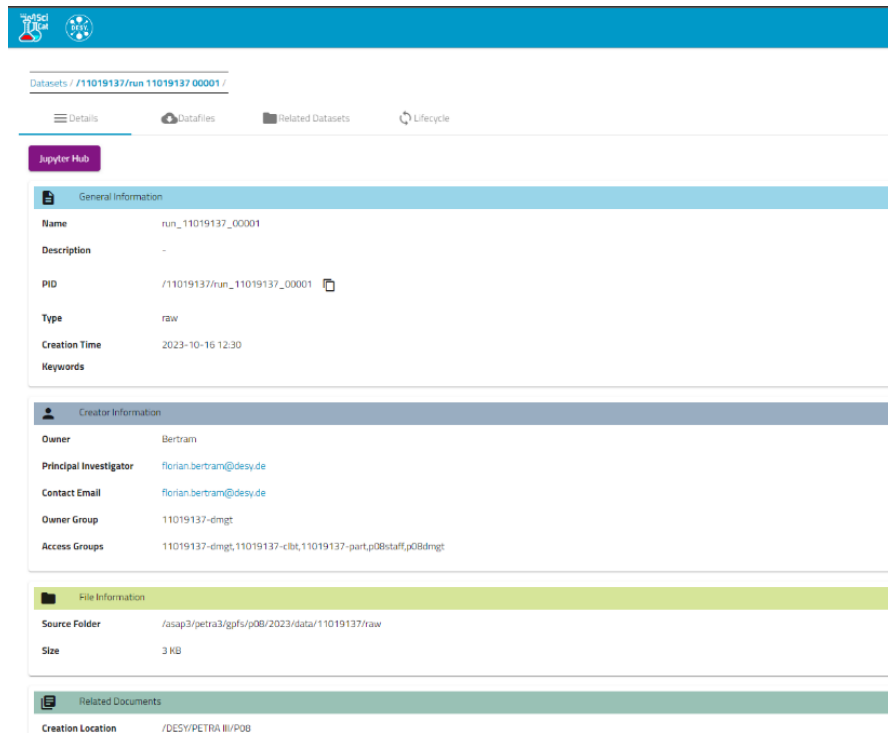


## What is SciCat?

- A scientific data cataloging system
- Designed to manage and publish experimental data from scientific research
- Supports FAIR principles (Findable, Accessible, Interoperable, Reusable)

## Purpose of SciCat:

- Facilitate data discovery and reuse
- Enhance collaboration among researchers
- Ensure long-term preservation of scientific data



The screenshot displays the SciCat web interface for a specific dataset. The top navigation bar is blue with the SciCat logo and a search bar. Below the navigation bar, the dataset path is shown: `Datasets / 11019137/run 11019137_00001 /`. The main content area is divided into several sections:

- General Information** (light blue header):
  - Name: `run_11019137_00001`
  - Description: `-`
  - PID: `/11019137/run_11019137_00001` (with a document icon)
  - Type: `raw`
  - Creation Time: `2023-10-16 12:30`
  - Keywords: (empty field)
- Creator Information** (light blue header):
  - Owner: `Bertram`
  - Principal Investigator: `florian.bertram@desy.de`
  - Contact Email: `florian.bertram@desy.de`
  - Owner Group: `11019137-dmgt`
  - Access Groups: `11019137-dmgt, 11019137-clbt, 11019137-part,p08staff,p08dmgt`
- File Information** (light green header):
  - Source Folder: `/asap3/petra3/gfbs/p08/2023/data/11019137/raw`
  - Size: `3 KB`
- Related Documents** (light green header):
  - Creation Location: `/DESY/PETRA III/P08`

## What is LinkML?

- A schema language for defining data models
- Supports creation of interoperable data structures
- Emphasizes semantic consistency and data integration



## Key Features of LinkML:

- YAML-based schema definitions
- Automatic generation of documentation, JSON schemas, and more
- Support for rich metadata annotations
- Additionally, it is a framework for working with and validating data in a variety of formats (JSON, RDF, TSV), with generators for compiling LinkML schemas to other frameworks and generating documentation.



Development of Data Model Preparation and Validation Tool for Scientific Metadata in SciCat.

**Primary Objectives:**

1. Data Model Preparation
2. Validation Layer Implementation
3. Documentation Generation
4. Metadata List Management

**Roadmap:****1. Centralized Git Repository:**

1. Centrally managed and curated repository for schema definitions.

**2. CI/CD Pipeline to Generate:**

1. Documentation (based on MKDoc)
2. JSON Schema
3. Simple spreadsheet-like view for discussion

**3. Foundation:**

1. Use results of DAPHNE TA1 and TA2 as the foundation for schema creation.

**4. Schema Expansion:**

1. Expand schemas to technique or beamline-specific use cases.

SciCat@DESY

Search docs

Index

BEAMLINES AND INSTRUMENTS

Class: P08Dataset

Class: P65Dataset

Class: FlashDatasetMetadataFile

PROJECTS

Class: Rockit

BASE CLASSES

Class: ScicatDataset

Class: DiffractionMetadata

TECHNIQUES

GIXD

UNITS

Class: Measurement

Class: AngleMeasurement

Class: EnergyMeasurement

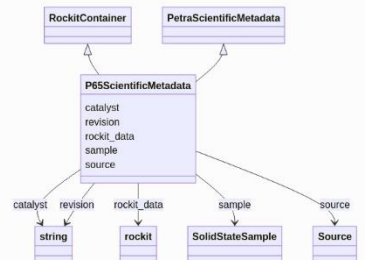
Class: LengthMeasurement

Class: TemperatureMeasurement

Class: P65ScientificMetadata

The metadata associated with P65

URI: [https://desy.de/linkml/opensource/beamline\\_p65/-P65ScientificMetadata](https://desy.de/linkml/opensource/beamline_p65/-P65ScientificMetadata)



Inheritance

- ScientificMetadataCommons
  - PetraScientificMetadata
    - P65ScientificMetadata (RockitContainer)

SciCat@DESY

Search docs

Index

BEAMLINES AND INSTRUMENTS

Class: P08Dataset

Class: P65Dataset

Inheritance

Slots

Identifier and Mapping Information

Mappings

LinkML Source

PROJECTS

Class: Rockit

BASE CLASSES

Class: ScicatDataset

Class: DiffractionMetadata

TECHNIQUES

GIXD

UNITS

Class: Measurement

Class: AngleMeasurement

Class: EnergyMeasurement

Class: LengthMeasurement

Class: TemperatureMeasurement

Beamlines and Instruments / Class: P65Dataset

Search docs

Index

BEAMLINES AND INSTRUMENTS

Class: P08Dataset

Class: P65Dataset

Inheritance

Slots

Identifier and Mapping Information

Mappings

LinkML Source

PROJECTS

Class: Rockit

BASE CLASSES

Class: ScicatDataset

Class: DiffractionMetadata

TECHNIQUES

GIXD

UNITS

Class: Measurement

Class: AngleMeasurement

Class: EnergyMeasurement


Class: LengthMeasurement

Class: TemperatureMeasurement

Class: P65Dataset

The metadata associated with P65

URI: [https://desy.de/linkml/opensource/beamline\\_p65/-P65Dataset](https://desy.de/linkml/opensource/beamline_p65/-P65Dataset)



Inheritance

- ScicatDataset
  - P65Dataset

Slots

Name	Cardinality and Range	Description
scientificMetadata	1..1 P65ScientificMetadata	
owner	1..1 xsd:string	Owner or custodian of the dataset, usually first name
ownerEmail	0..1 xsd:string	Email of the owner or custodian of the dataset
orcidOfOwner	0..1 xsd:string	ORCID of the owner or custodian
contactEmail	0..1 xsd:string	Email of the contact person for this dataset
sourceFolder	0..1 xsd:string	Absolute file path on file server containing the files of

- Gitlab datamodel link - <https://gitlab.desy.de/fs-ec/scicat/datamodel>
- LinkML Documentation: <https://linkml.io/linkml/>
- Data Model Documentation: <https://fs-ec.pages.desy.de/scicat/datamodel/>

Through the use of LinkML we aim at providing:

1. The possibility to automatically convert a dataset in the DESY-SciCat into content that is ingestible by external tools through the use of ontologies (e.g. FAIRmat's NOMAD, NFDI4Cat, NFDI4Chem, ISPyB )
2. Provide validator service (ideally backed into SciCat backend) that provides information on whether or not a certain dataset complies with the proposed schema
3. Empower scientists to participate in the schema definition process

A json schema for SciCat's ScientificMetadata that is build of standardized blocks, but customized for each beamline

*Your input is welcome!*

*These are just first steps, nothing caved in stone yet!*

*We are open for advice!*

**Thank you for your attention !!**