

Scitacean

User-oriented SciCat client for Python

PRESENTED BY JAN-LUKAS WYNEN

2025-06-25



Sci... what?

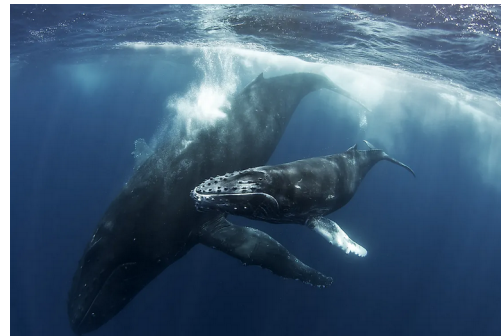
- Say 'citation'

science



- Shutterstock

+ cetacean



- Martin Prochazkacz

- scicatproject.org/scitacean
- Python client for SciCat for scientists



Why a new tool?

REST API



Dataset

Metadata:

- Source folder
- Size
- Owner group
- ...

Files



- Actual files
- Directory structure
- Separate access
 - SFTP
 - Direct file access
 - Globus
 - ...

Data File

- Path
- Size
- Checksum
- ...



Downloading data

```
from scitacean import Client

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token("ess", token=token)

pid = "20.500.12269/1b7c5c6e-dd03-453a-bfe6-4d07e9d86c77"
dataset = client.get_dataset(pid)
dataset = client.download_files(dataset, target="./data")

dataset
```

Downloading data

```
from scitacean import Client

token = "eyJhbGciOiJIUzI1NiIsI..."  
client = Client.from_token("ess", token)  
  
pid = "20.500.12269/1b7c5c6e-dd03-453"  
dataset = client.get_dataset(pid)  
dataset = client.download_files(datas  
  
dataset
```

User /

User Information



Name Jan-Lukas

Email Jan-Lukas.Wynen@ess.eu

Id

Groups

SciCat Token eyJhbGciOiJIUzI1NiIsI... 

Auth strategy ldap

Settings

Frontend Release Version local.dev

Site ESS



Downloading data

```
from scitacean import Client

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token("ess", token)

pid = "20.500.12269/1b7c5c6e-dd03-453a-
dataset = client.get_dataset(pid)
dataset = client.download_files(dataset)

dataset
```

SciCat EUROPEAN SPALLATION SOURCE STAGING

Datasets / 20.500.12269/1b7c5c6e-dd03-453a-bfe6-4d07e9d86c77 /

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

General Information

Dataset Name	Simulated silicon powder sample
Description	McStas + GEANT4 simulation of DREAM in high-flux configuration with a silicon sample. data_dream_hf_si_bc_240_wlgh.csv.zip : Detector data. - Cave_TOF_Monitor_hf_si_bc_240.dat : Cave monitor intensity vs ToF. - Cave_BeamLambda_monitor_hf_si_bc_240.dat : Cave monitor wavelength.
Pid	20.500.12269/1b7c5c6e-dd03-453a-bfe6-4d07e9d86c77
Type	raw
Creation time	2025-02-05 10:23
Keywords	DREAM McStas simulation high flux GEANT4



Downloading data

```
from scitacean import Client

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token("ess", token=token)

pid = "20.500.12269/1b7c5c6e-dd03-453a-bfe6-4d07e9d86c77"
dataset = client.get_dataset(pid)
dataset = client.download_files(dataset, target="./data")

dataset
```

RawDataset

	Name	Type	Value	Description
*	creation_time	datetime	2025-02-05 09:23:48+0000	Time when dataset became full...
*	input_datasets	list[PID]	[]	Array of input dataset identifier...
*	source_folder	RemotePath	RemotePath('/ess/data/dream/2025/037980/u...	Absolute file path on file server...
	description	str	McStas + GEANT4 simulation of DREAM in hig...	Free text explanation of conten...
	name	str	Simulated silicon powder sample	A name for the dataset, given b...
🔒	pid	PID	20.500.12269/1b7c5c6e-dd03-453a-bfe6-4d07...	Persistent identifier of the data...
	proposal_id	str	037980	The ID of the proposal to which...



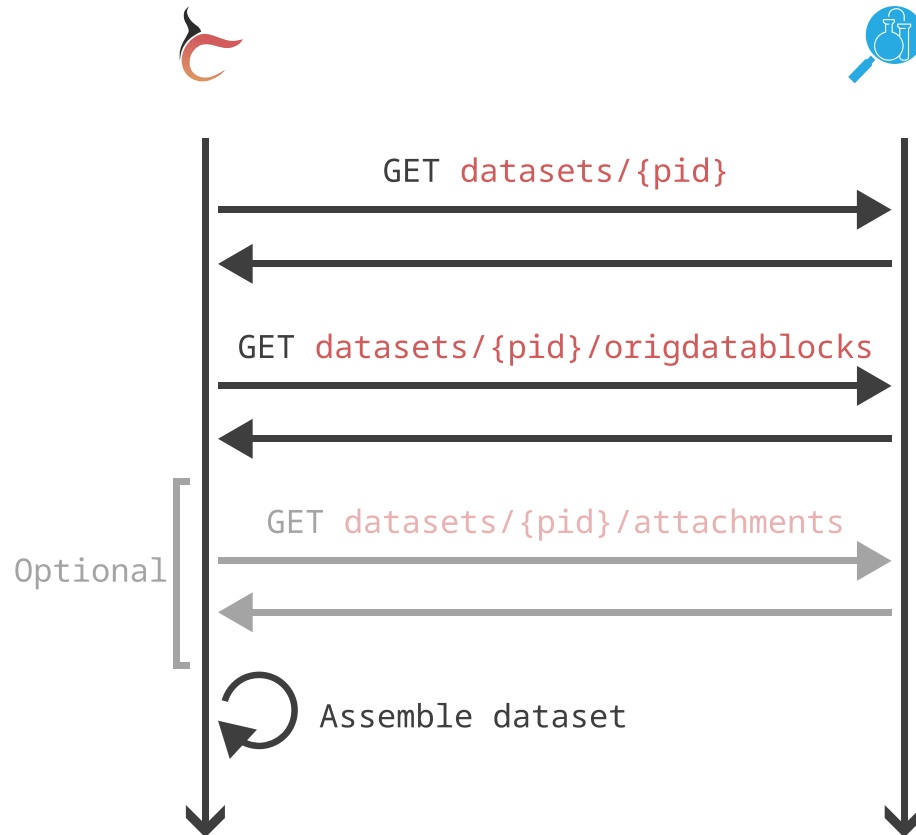
Uploading data

```
from scitacean import Client, Dataset

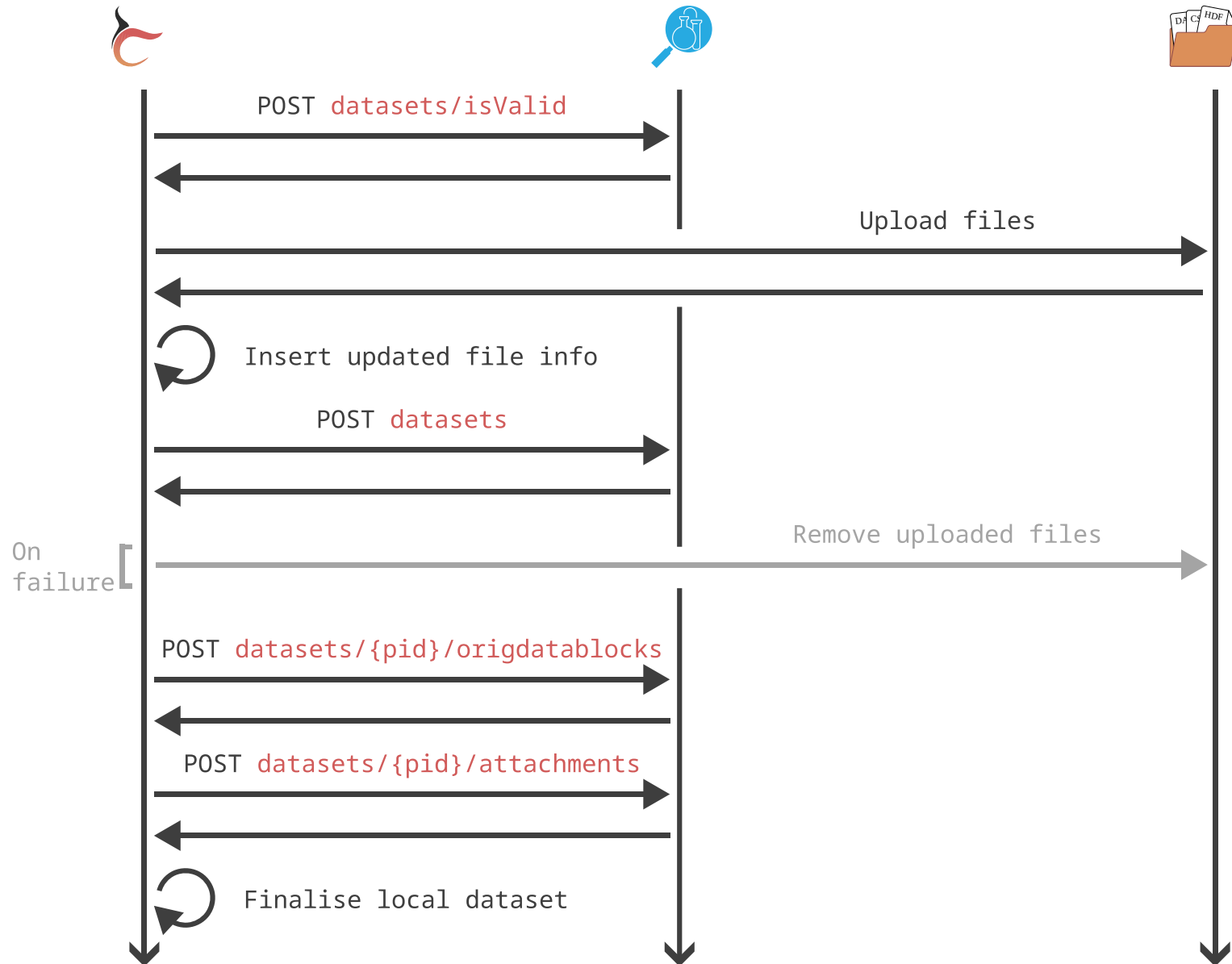
dataset = Dataset(
    name="My dataset",
    owner="Jan-Lukas Wynen",
    source_folder="/data/my_dataset",
    ...
)
dataset.add_files("file1.csv", "file2.dat")

token = "eyJhbGciOiJIUzI1NiIsI...
client = Client.from_token("ess", token=token)
client.upload_new_dataset_now(dataset)
```

```
dataset = client.get_dataset(pid)
```



```
finalised = client.upload_new_dataset_now(dataset)
```





File transfer

Different file [storage & access](#):

- Direct access
- Globus
- SSH / SFTP
- etc.

```
from scitacean.transfer.copy import CopyFileTransfer
from scitacean.transfer.link import LinkFileTransfer
from scitacean.transfer.select import SelectFileTransfer
from scitacean.transfer.sftp import SFTPFileTransfer

# At ESS:
transfer = SelectFileTransfer([
    LinkFileTransfer(),
    CopyFileTransfer(),
    SFTPFileTransfer(host="login.esss.dk"),
])
```



Future work

- Authentication
 - Unified SciCat + file server (specific to facility)
 - Interactive login
 - Persistent login
- Querying
 - Need high-level interface

```
client.scicat.query_datasets(  
    {"proposalIds": ["bc.123"]},  
    limit=5,  
    order="creationTime:desc",  
)
```



Scitacean

Pip:

```
pip install "scitacean[sftp]"
```

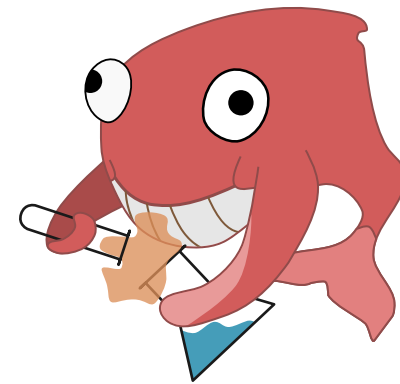
Conda:

```
conda install -c conda-forge scitacean
```

Documentation:

scicatproject.org/scitacean

Remember: Science Whales!



Thank you for your attention



Connecting to SciCat

```
from scitacean import Client

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token("ess", token=token)
```

```
from scitacean import Client
from scitacean.transfer.sftp import SFTPFileTransfer

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token(
    url="https://scicat.ess.eu/api/v3",
    token=token,
    tile_transfer=SFTPFileTransfer(host="login.esss.dk")
)
```



Authentication (SciCat)

```
from scitacean import Client
token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token("ess", token=token)
```

```
from scitacean import Client
client = Client.from_credentials(
    "ess",
    username="janlukaswynen",
    password="..."
)
```

```
from scitacean import Client
client = Client.without_login("ess")
```



Authentication (SFTP)

```
from scitacean import Client
from scitacean.transfer.sftp import SFTPFileTransfer

token = "eyJhbGciOiJIUzI1NiIsI...\"
client = Client.from_token(
    \"ess\",
    token=token,
    transfer=SFTPFileTransfer(host=\"login.esss.dk\"),
)
```

```
transfer = SFTPFileTransfer(
    host=\"login.esss.dk\",
    username=\"janlukaswynen\",
    password=\"...\",
)
```



Low-level interface

```
from scitacean import Client

client = Client.from_token("ess", token=token)

ds = client.scicat.get_dataset_model(pid)
db = client.scicat.get_orig_datablocks(pid)

client.scicat.create_dataset_model(ds)
client.scicat.create_orig_datablock(db, dataset_id=ds.pid)
```

Similar to [pyscicat](#)

```
from scitacean.transfer.sftp import SFTPFileTransfer

transfer = SFTPFileTransfer(host="login.esss.dk")
with transfer.connect_for_download(<...>) as con:
    con.download_files(
        remote=["/remote/file.dat"],
        local=["./local-data/file.dat"],
    )
```