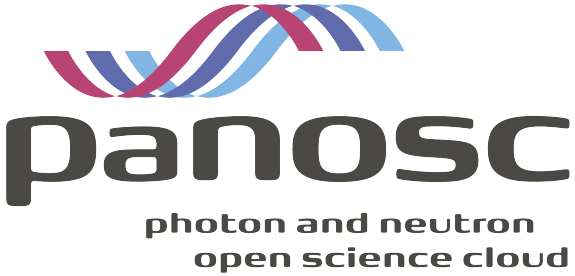


Search API

Gareth Murphy
Data Curation Scientist
European Spallation Source

A Tale of Two APIs

- PaNOSC WP3 has two main functions
  1. Provide our open data to European Open Science Cloud
  2. Provide data (embargoed and open) catalogue to PI and public

EOSCpilot
The European Open Science Cloud for Research Pilot Project

Harvest API

- Open metadata, no login/password
- 3 year old, unembargoed or published data
- XML description
- Based on pre-existing OAI-PMH open standard for metadata
- Connects to OpenAire, B2FIND

Search API

- Login/password with ORCiD/eduGAIN required for embargoed
- Recent data, embargoed and permission/access controlled
- JSON description
- Access to data download and/or online Jupyter/analysis notebook
- Newly developed API with input from ExPaNDS and PaNOSC community

PaNOSC search

Filters

Data Type

- Simulation 521
- Experiment 2560
- Derived 423

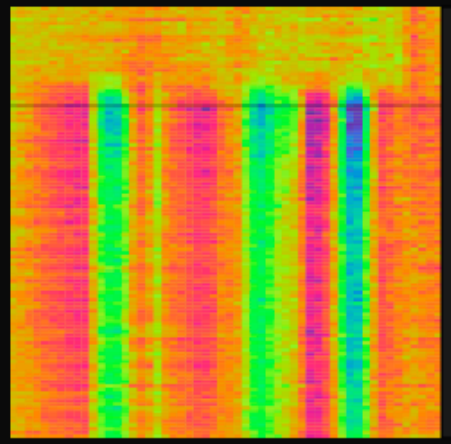
Field

- X-Ray Sources 368
- Plasma Physics 49
- Ion Acceleration 76
- Electron Acceleration 85
- Material and Biomolecular Applications 122

Technique

- X-ray phase contrast imaging 59
- X-ray Diffraction 45
- X-ray absorption spectroscopy 85
- Coherent Diffractive Imaging 26
- Atomic, Molecular and Optical Science 736
- Soft X-ray Materials Science 48

Datasets



Time-resolvent spectroscopy - run 1-52

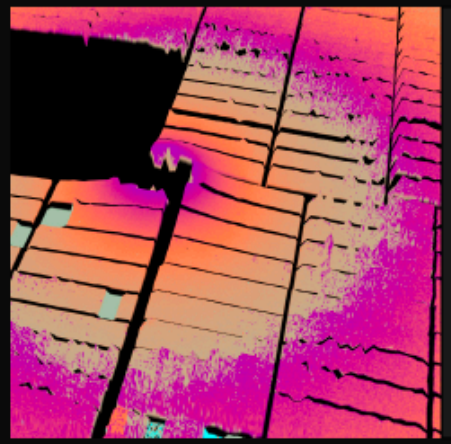
RP4-SRS focuses on time-resolvent spectroscopy experiments in the full range of frequencies from IR to UV. Users can measure samples as varied as solid state crystals, or proteins in their natural environment. Time-resolved spectroscopy is the collection of techniques that are used to examined the dynamic processes of materials and chemicals upon illumination with a pulsed laser...

Dataset | X-ray Spectroscopy | Pulsed Radiolysis | All Tags 8

Created 2019/03/15
Size 328 MB
Views 3

jupyterlab

launch VM



Two-color XUV+NIR femtosecond photoionization of neon in the near-threshold region

RP4-SRS focuses on time-resolvent spectroscopy experiments in the full range of frequencies from IR to UV. Users can measure samples as varied as solid state crystals, or proteins in their natural environment. Time-resolved spectroscopy is the collection of techniques that are used to examined the dynamic processes...

Dataset | X-ray Spectroscopy | XFEL

Created 2019/03/15
Size 7 GB
Views 3

jupyterlab

launch VM

Laser-driven Ion Acceleration from Plastic Target

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt

Created 2021/11/03

Some time pressure …

- The API definition is a deliverable for end of May 2020 (Month 18)
- 3 months - not much time for major changes to the data model
- But we value your opinion …

The relationship of these work packages is shown in figure below.

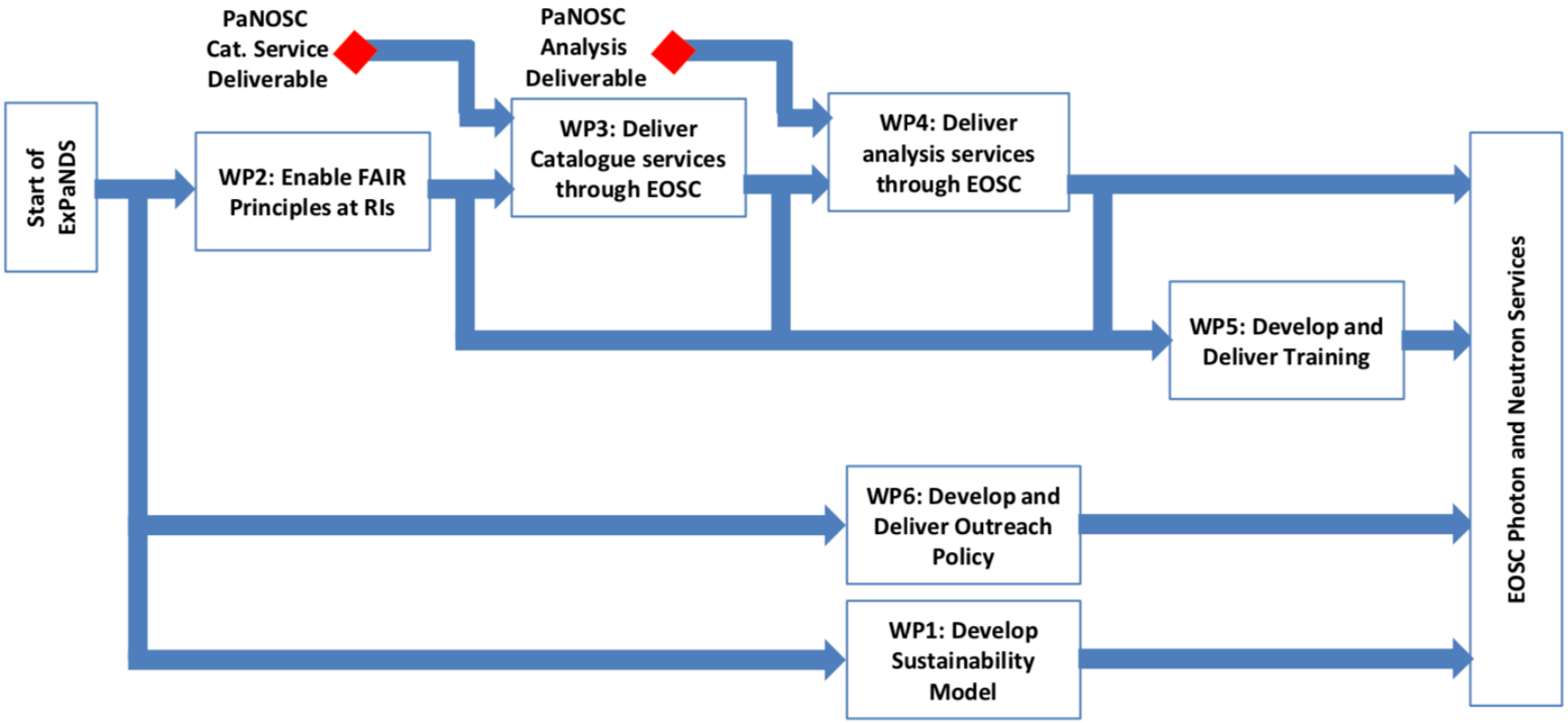


Figure 3.1.5: Relationship between ExPaNDS work packages

Scientific Use Cases

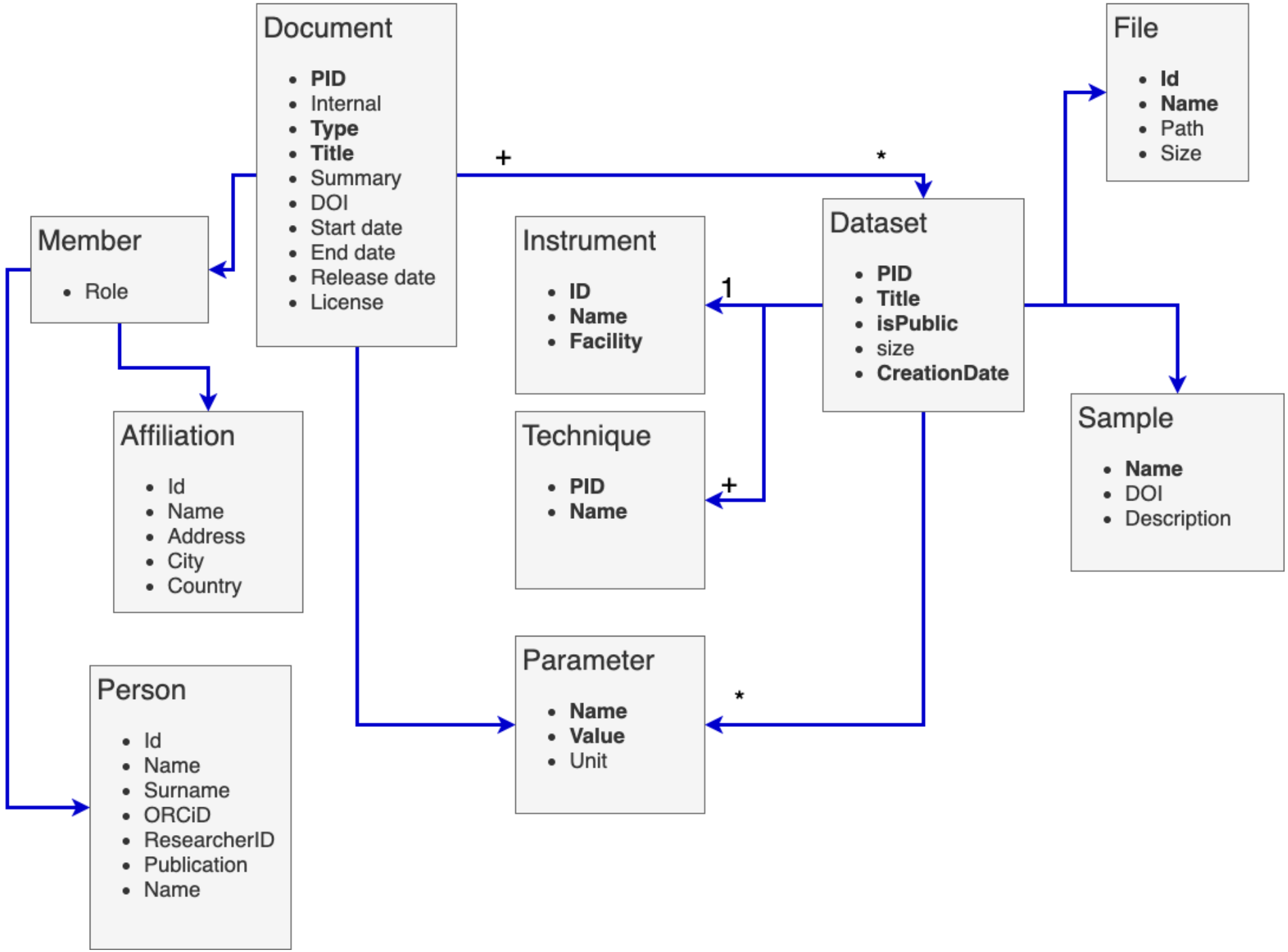
Actor	Story
Scientist	I would like to know the states of Machine (i.e Energy) during an experiment
PTM	I would like to search for all metadata produced from the experiment (note that this are application specific)
Scientist	I would like to know the type of the software used for the analysis of data
Scientist	I want to be able to search for all data on water at the PanOSC facilities
Scientist	I would like to search for experiments/datasets by a particular metadata entity [eg sample, formula, title, scientific technique, sample type]
Scientist	I would like to search for detector type
Scientist	I would like to know the instruments setup
Scientist	I would like to retrieve the logbook of an experiment
Scientist	I would like to find all experimental data associated with a proposalId or experimental number
Scientist	I would like to find all experimental data matching abstract/proposal text
Scientist	I would like to see the proposal details associated with experimental data (abstract, title, experimental dates)
Scientist	I would like to search for data from a particular institute / instrument
Scientist	I would like to search for data acquired during a particular period
PTM	I would like to find my own data (experimental data from my proposals)

Actors

Actor	Description
Proposal Team Member (PTM)	Person involved in a proposal
Open Access User (OAU)	Public member wanting to analyse open data
Scientist	Regroups Proposal Team Member and Open Access User
Admin	IT Staff
Facility Management (FM)	Directors, PaNOSC stakeholders (requiring metrics, usage, KPI...)
WP4 (Data Analysis Services)	Uses WP3 API to obtain metadata and to obtain access to data

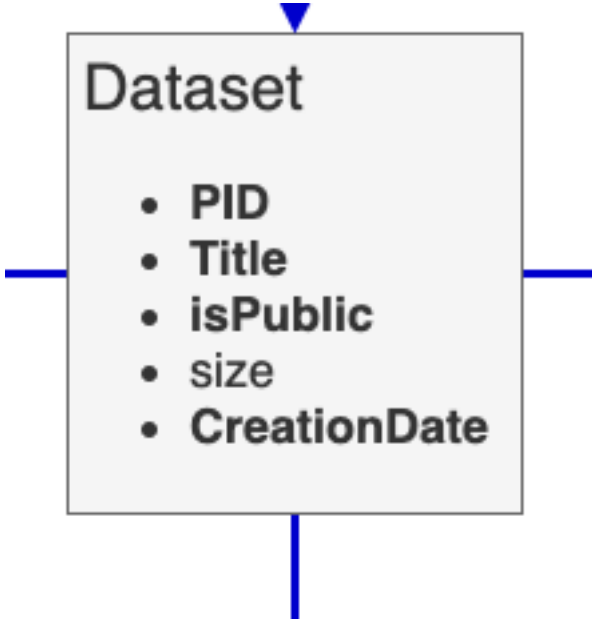
WP4 Use Cases

Actor	Story	
WP4	For the authenticated user, I would like to obtain a list (full metadata) of their associated proposals or data from different institutes	
WP4	I would like to obtain a list of open data (restricted metadata) from different institutes using specific search criteria	
WP4	Having identified a proposal or datafile/set of interest I would like to obtain the data in order to to analyse it	Tl ol m er


Document
PID
Internal
Type
Title
Summary
DOI
Start date
End date
Release date
License
File
Id
Name
Path
Size
Member
Role
Instrument
ID
Name
Facility
Dataset
PID
Title
isPublic
size
CreationDate
Affiliation
Id
Name
Address
City
Country
Technique
PID
Name
Sample
Name
DOI
Description
Person
Id
Name
Surname
ORCiD
ResearcherID
Publication
Name
Parameter
Name
Value
Unit
+
*
1
+
*

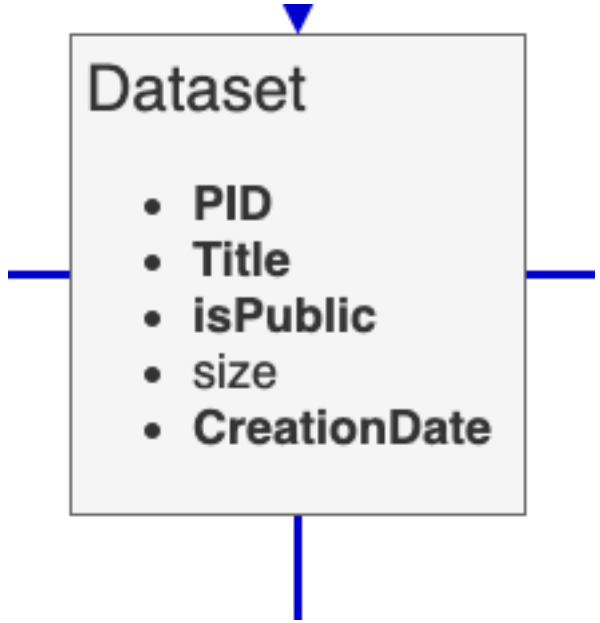
Dataset

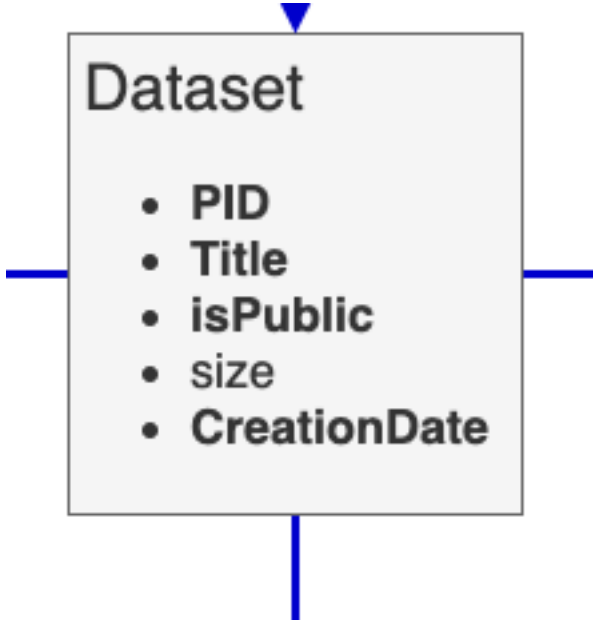
- PID - persistent identifier (DOI, Handle)
- Title - string
- isPublic - boolean
- Size - optional
- Creation date



Example


```
[
  {
    "pid": "20.500.12269/2d5af6ef-6b94-43a4-972b-4bdb4f6c95a7",
    "isPublic": true,
    "title": "V20 sample data",
    "creationDate": "2020-02-06T13:02:20.000Z",
    "size": 0,
    "parameters": [
      {
        "name": "wavelength",
        "value": 1e-11,
        "unit": "m"
      },
      {
        "name": "sample_temperature",
        "value": 300,
        "unit": "degK"
      }
    ]
  }
]
```



SciCat | ESS European Spallation Source

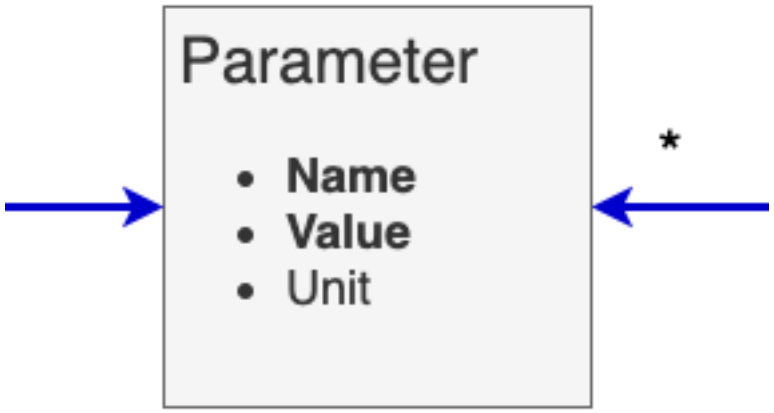
Jupyter Hub

About the data

Name	Last Neutrons Ever at HZB.
Description	V20 data
Owner	Peter Kadletz
PID	20.500.12269/2511nicos_00002511.hdf
Source Folder	/nfs/groups/beamlines/v20/YC7SZ5
Keywords	v20 ⊗ neutron ⊗ Add Keyword +

Structural information

Type	raw
Version	3.0.1
Proposal	Final prototype
Sample	Sample for V20
Size	203 MB
Orcid	default



Administrative information

Creation Time	2020-02-06 14:02
Owner Group	ess
Access Groups	
Contact Email	gareth.murphy@esss.se

Derived Data

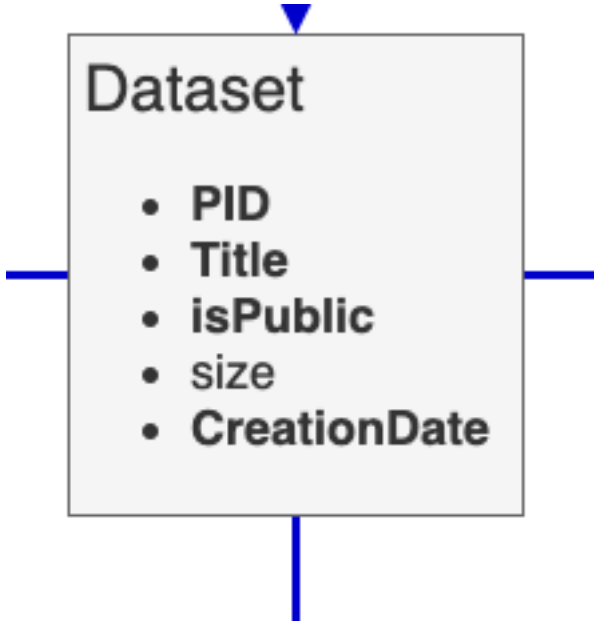
Investigator	gareth.murphy@esss.se
Input Datasets	
Software Used	mantid

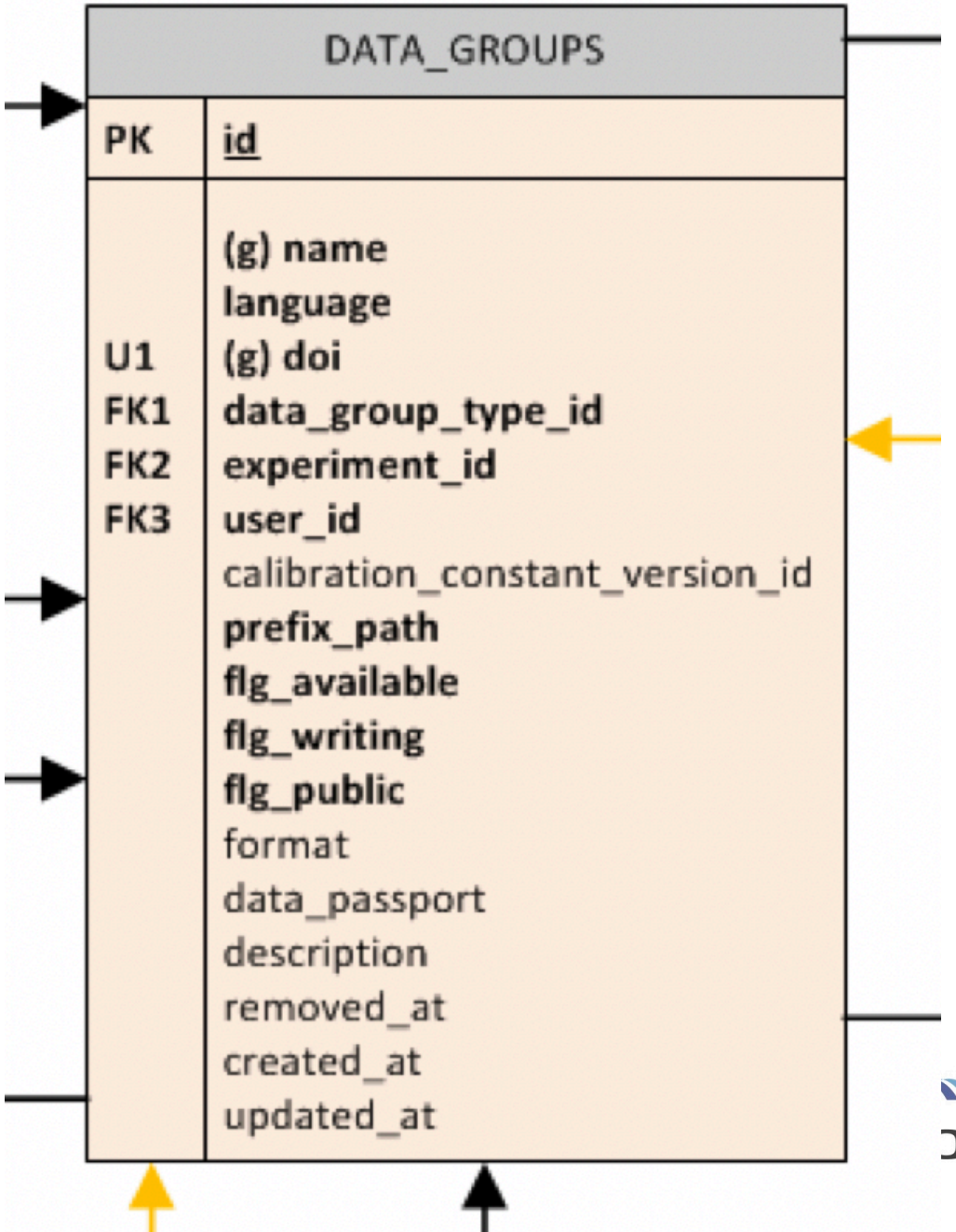
Scientific Metadata

View | Edit

Name	Value	Unit
Wavelength	1e-11	m
Sample Temperature	300	degK

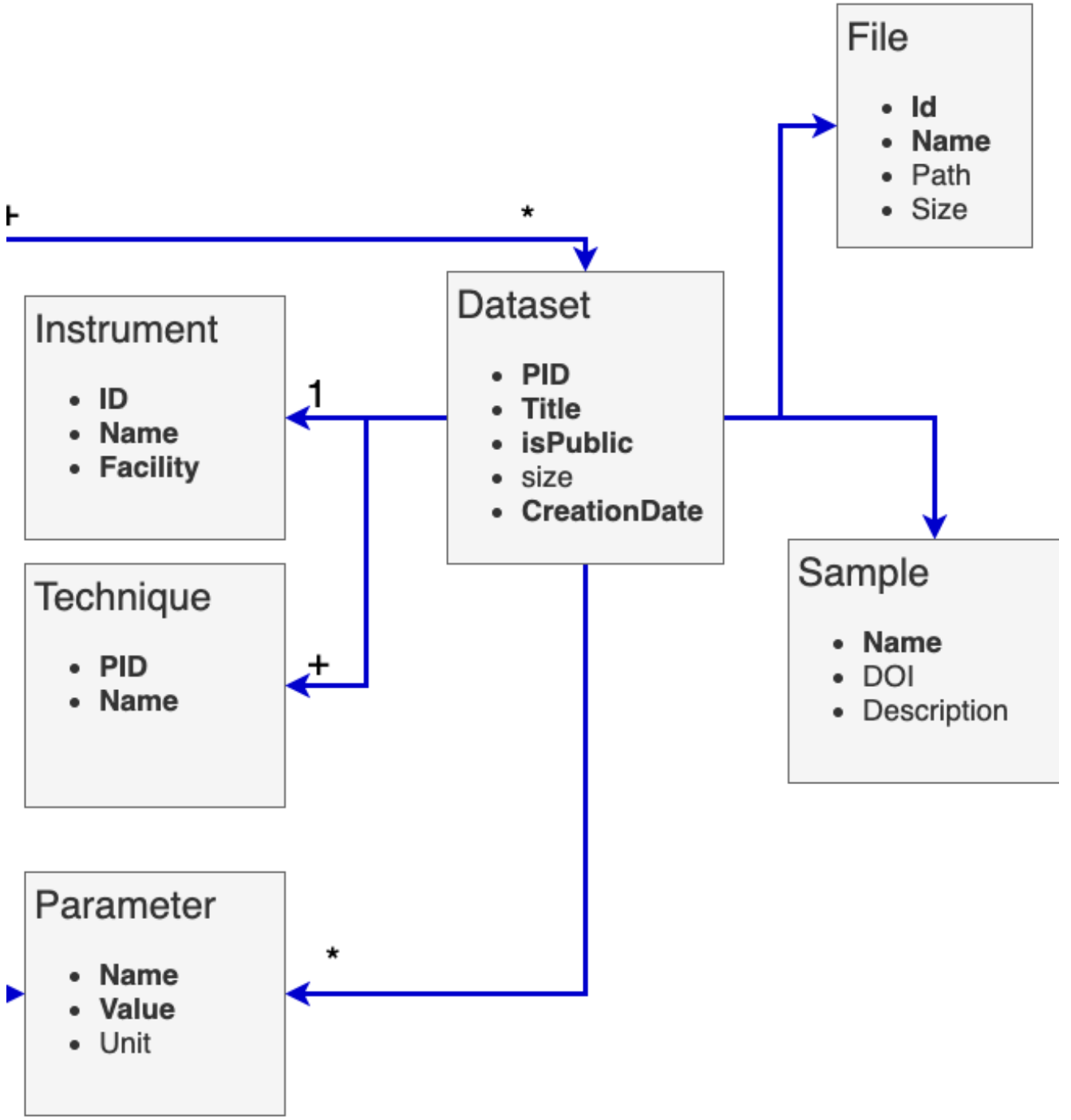
PaN -> MyMdc


Dataset
PID
Title
isPublic
size
CreationDate


DATA_GROUPS
PK
id
(g) name
language
U1
(g) doi
FK1
data_group_type_id
FK2
experiment_id
FK3
user_id
calibration_constant_version_id
prefix_path
flg_available
flg_writing
flg_public
format
data_passport
description
removed_at
created_at
updated_at

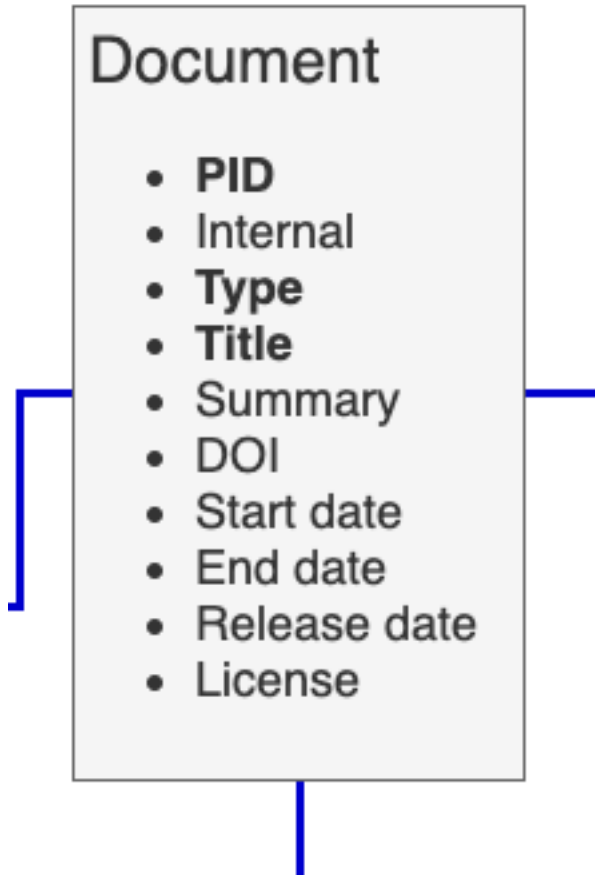
Dataset (II)

- File(s) - where actual data is
- Sample(s) - material examined
- Instrument(s) - beam line
- Technique - from ontology e.g. small angle scattering (SANS)
- Parameter - e.g. Wavelength 1 Angstrom



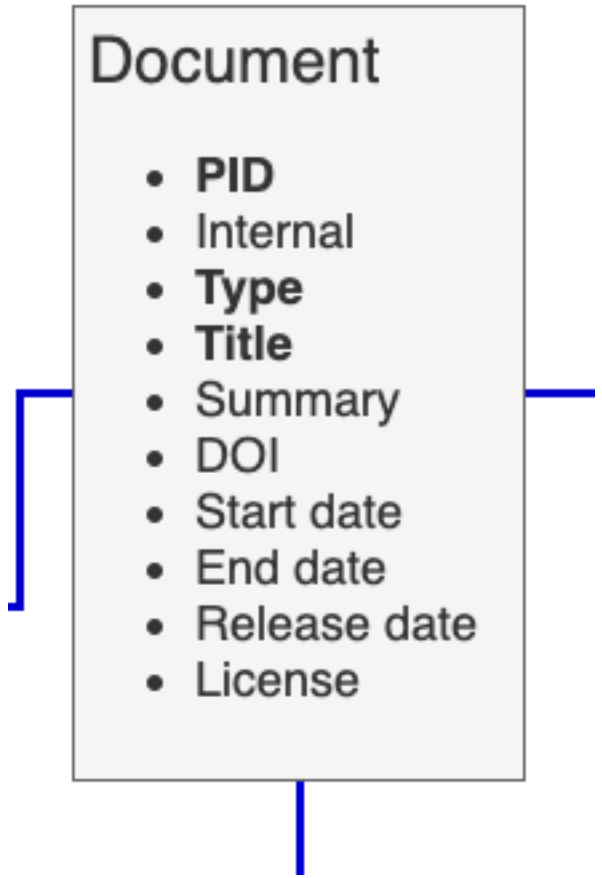
Document

- Can be Proposal or Publication/ PublishedDatasets or Other
- Type = {Proposal,Publication}
- Title
- Internal - boolean yes/no
- Summary - description
- Dates, start, end, or beam time
- Release data, when public
- License - CC-BY-4.0 or ?



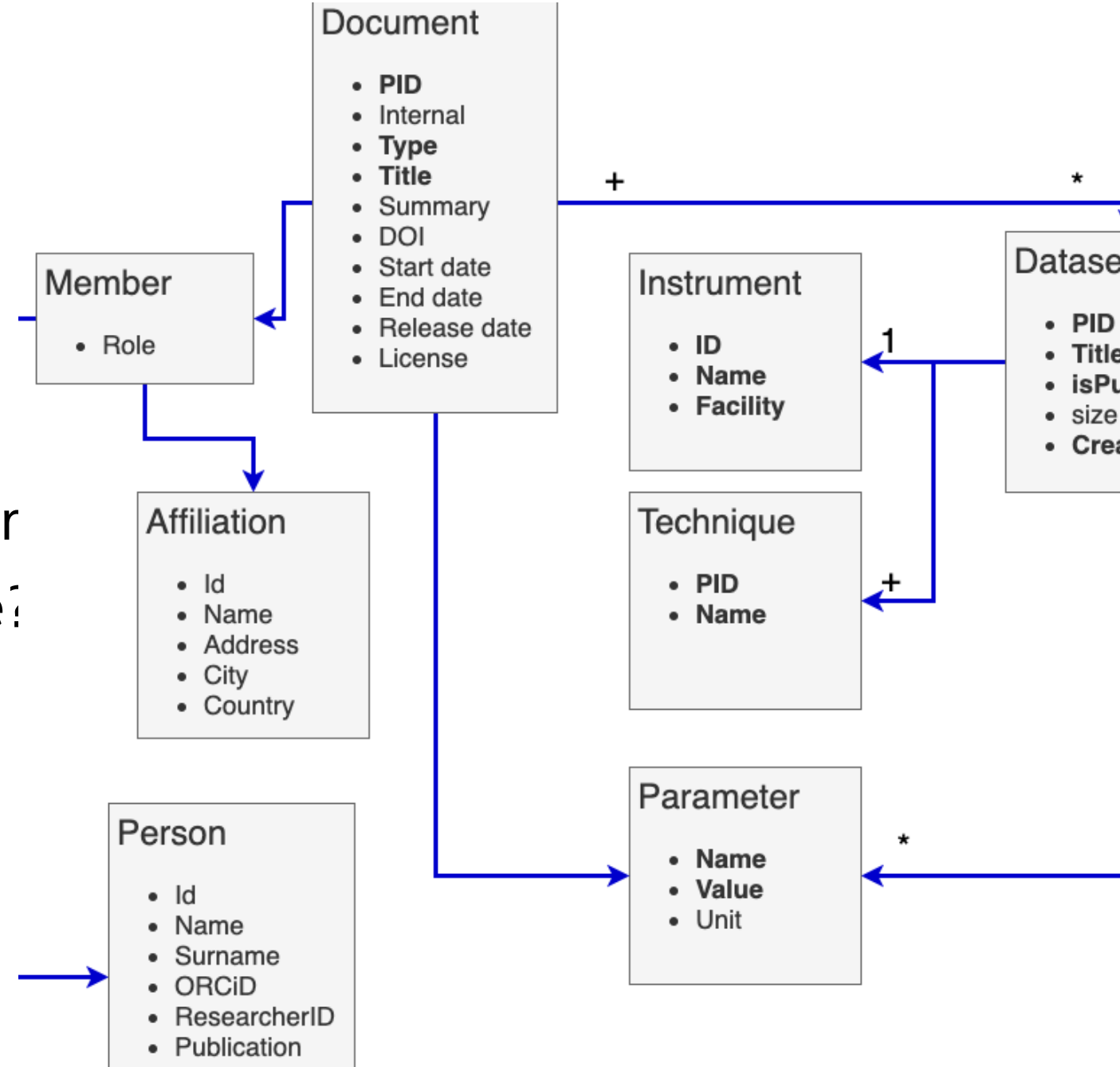
Example Document

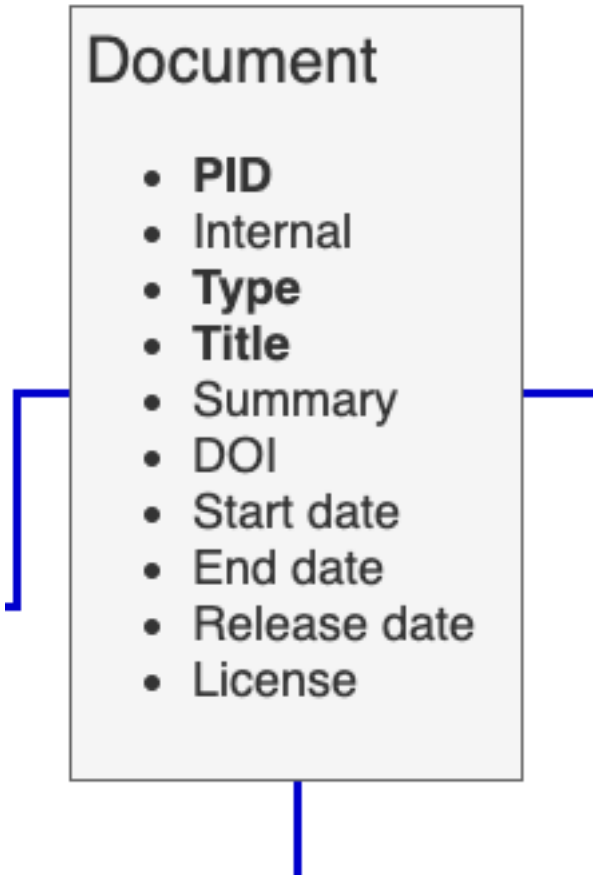

```
{
  "pid": "10.17199/BRIGHTNESS/V200161",
  "title": "Sample Data from V20",
  "internal": true,
  "summary": "String",
  "type": "String",
  "startDate": "2020-02-02",
  "endDate": "2020-02-02",
  "releaseDate": "2020-02-02",
  "license": "CC-BY-4.0"
}
```

Document II

- Datasets - generated in proposal or curated by the author
- Member with Role - principal investigator or co-investigator
- Affiliation - can have multiple?
- Person - human being
- Parameter - summary, average, max of what is in Dataset





Published Datasets / 10.17199/NXMV08.DSC0001 /

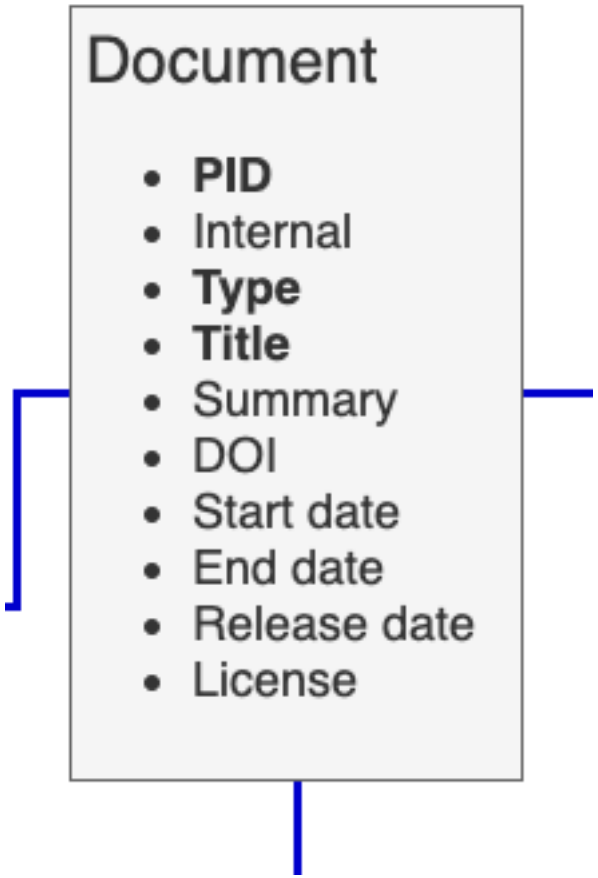
About the published data

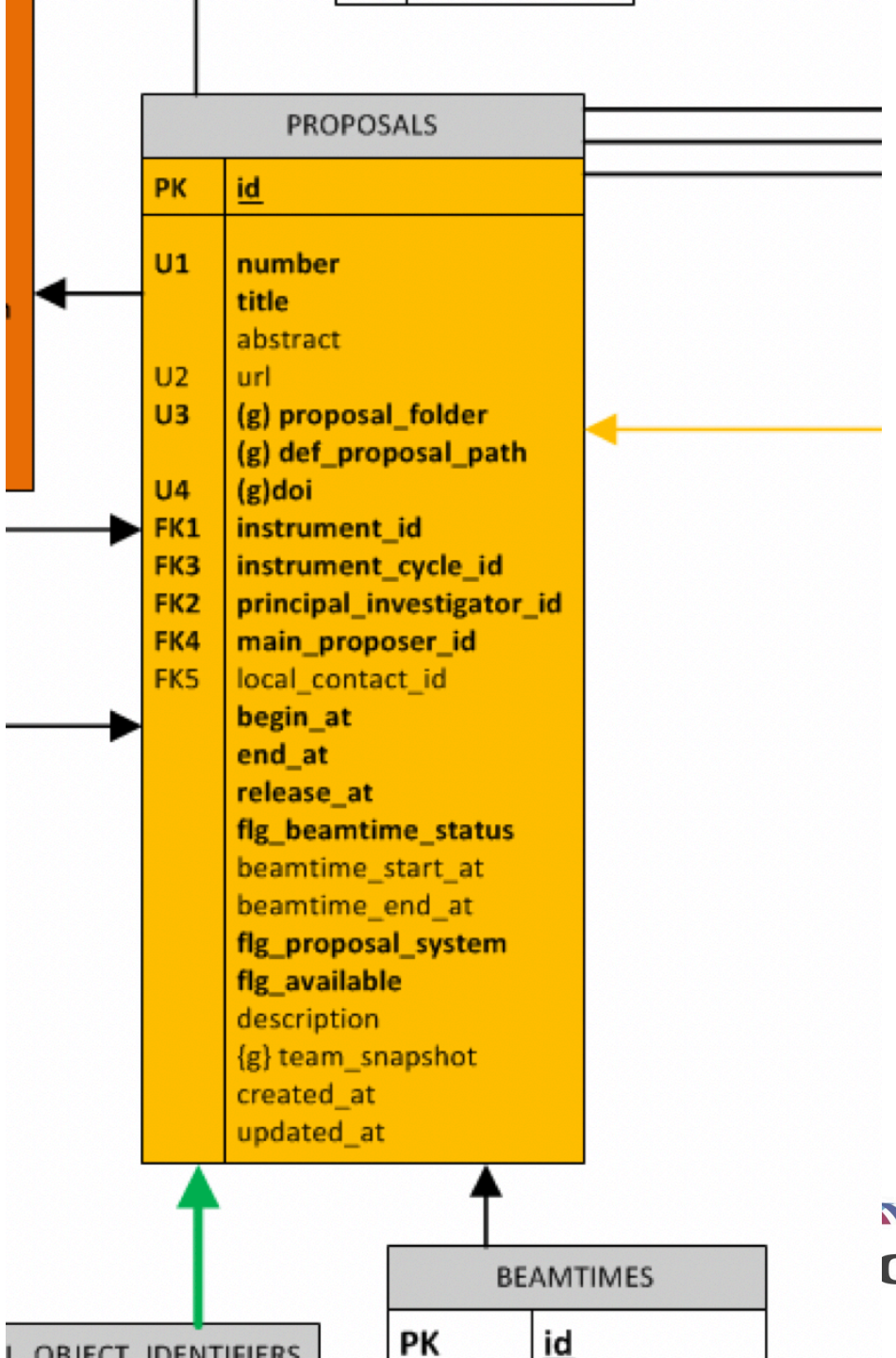
Title	Differential scanning calorimetry (DSC) data for breast cancer cells
Creator List	Heloisa Nunes Bordallo,Murillo Longo Martins
Abstract	Datasets from differential scanning calorimetry (DSC) data for breast cance
DOI	10.17199/NXMV08.DSC0001
URL	doi.esss.se/detail/10.17199/NXMV08.DSC0001

Administrative metadata

Scicat User	ingestor
Creator	Heloisa Nunes Bordallo,Murillo Longo Martins
Size	1 MB
Number of Files	5
Data Description	github.com/ess-dmsc/ess_file_formats/wiki/DSC
Dataset IDs	20.500.12269/20.500.12269/NXMV08.0001
Affiliation	ESS
Publisher	ESS

PaN -> MyMdC


Document
PID
Internal
Type
Title
Summary
DOI
Start date
End date
Release date
License


PROPOSALS
PK id
U1 number
title
abstract
U2 url
U3 (g) proposal_folder
(g) def_proposal_path
U4 (g)doi
FK1 instrument_id
FK3 instrument_cycle_id
FK2 principal_investigator_id
FK4 main_proposer_id
FK5 local_contact_id
begin_at
end_at
release_at
flg_beamtime_status
beamtime_start_at
beamtime_end_at
flg_proposal_system
flg_available
description
{g} team_snapshot
created_at
updated_at
BEAMTIMES
PK id

LoopBack Application 1.0.0 OAS3

./openapi.json

Servers

http://localhost:3000

Filter by tag

DatasetController

GET /datasets/{id}/metadata

GET /datasets/{id}

GET /datasets

DocumentController

GET /documents/count

GET /documents/{id}

GET /documents

InstrumentController

GET /instruments/count

GET /instruments/{id}

GET /instruments

SampleController

GET /samples/count

GET /samples/{id}

GET /samples

Unit Support

- Query to get range over units
- Loopback-like syntax
- { where : { variable: pressure, operator: gt, value: 1 , unit:bar } }
- Return all values where pressure is great than one bar
- Unit conversion should be handled.
- Need to know units at the other end

panosc

Connecting to your home facility

- External connection abstracted as a PanService
- When you connect your catalogue, add a new PanServiceProvider

Converting dataset

```
199  }
200
201  export function convertToPaN(scicatDataset: SciCatDataset) {
202    const panDataset: PanDataset = {
203      pid: scicatDataset.pid,
204      isPublic: true,        You, 4 days ago • move scicat specific code out of cont
205      title: scicatDataset.datasetName,
206      creationDate: scicatDataset.creationTime,
207      size: scicatDataset.size,
208    };
209    const paramArray: Measurement[] = [];
210    if ('scientificMetadata' in scicatDataset) {
211      Object.keys(scicatDataset.scientificMetadata).forEach((key: string) => {
212        // console.log('key', key);
213        const panParam = {
214          name: key,
215          value: scicatDataset.scientificMetadata[key]['value'],
216          unit: scicatDataset.scientificMetadata[key]['unit'],
217        };
218        paramArray.push(panParam);
219      });
220      panDataset.parameters = paramArray;
221    }
222    return panDataset;
```

panosc

Converting units

```
export function convertUnits(name: string, value: number, unit: string) {
  const qtyString = String(value) + ' ' + unit;
  const qty = new Qty(qtyString);
  const convertedQuantity = qty.toBase().toString();

  const convertedUnit = convertedQuantity.substr(
    convertedQuantity.indexOf(' ') + 1,
  );
  const convertedValue = convertedQuantity.substr(
    0,
    convertedQuantity.indexOf(' '),
  );
  const floatConverted = parseFloat(convertedValue);
  // add logic for wavlength in units of energy
  if (name === 'wavelength' && convertedUnit === 'J') {
    // if units are in energy
    // convert to joules than length
    const planckConstant = 6.64e-34;
    const speedOfLight = 3e8;
    const lambda = (planckConstant * speedOfLight) / floatConverted;
    return lambda;
  }
  return floatConverted;
```

panosc