SciCat at DESY

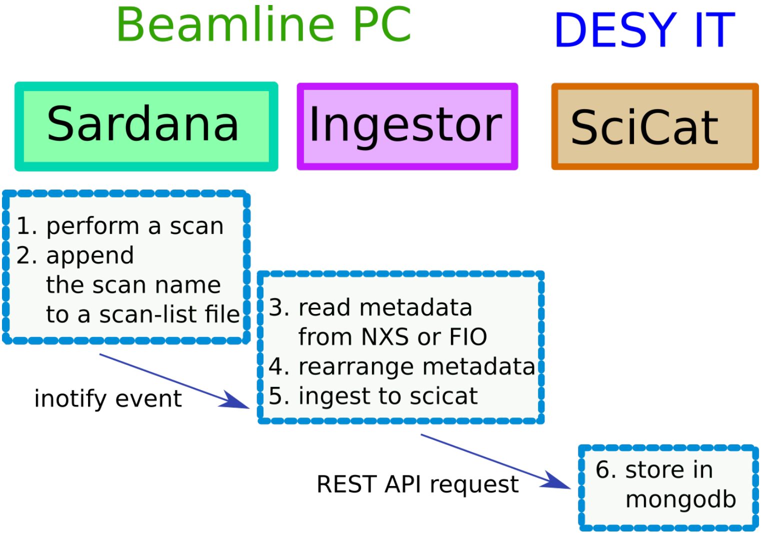
Grouping SciCat Scan datasets in Measurement datasets

Jan Kotański

Deutsches Elektronen-Synchrotron

February 12, 2024

Dataset ingestor



Beamline PC

DESY IT

Sardana

Ingestor

SciCat

1. perform a scan
2. append the scan name to a scan-list file

inotify event

3. read metadata from NXS or FIO
4. rearrange metadata
5. ingest to scicat

REST API request

6. store in mongodb

The scan-list file may also contain start/stop measurement tags, i.e. `__command__`

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Scan-list file example

```
myscan_00012
myscan_00013

__command__ start my_measurement
myscan_00014
myscan_00015
myscan_00016
__command__ stop
my_measurement:1704367280.9590826

myscan_00017

__command__ start my_measurement
myscan_00017:1704367282.023423
__command__ stop
my_measurement:1704367286.3453534
```

Rearranging scientificMetadata with copy-map

The scientificMetadata can be changed i.e. copied or removed, by copy-map {target: source}

- a file containing the copy-map is defined by **metadata_copy_map_file** *(str)* , default: `None`
  `/home/p08user/.config/DESY/scingestor-metadata-copy-map.lst`
- the copy-map file contain `json`, `yaml` or `text` file with a dictionary or two column array

- the copy-map can be also defined in a NeXus field: `/:NXentry/nxsfileinfo_parameters:NXparameters /copymap`
- the above NeXus field copy-map would be provided by a NeXus component
  so different copy-maps could be used in various measurement groups

Rearranging scientificMetadata with copy-map

The scientificMetadata can be changed i.e. copied or removed, by copy-map {target: source}

- a file containing the copy-map is defined by **metadata_copy_map_file** *(str)* , default: `None`
  `/home/p08user/.config/DESY/scingestor-metadata-copy-map.lst`
- the copy-map file contain `json`, `yaml` or `text` file with a dictionary or two column array

- the copy-map can be also defined in a NeXus field: `/:NXentry/nxsfileinfo_parameters:NXparameters /copymap`
- the above NeXus field copy-map would be provided by a NeXus component
  so different copy-maps could be used in various measurement groups

ScientificMetadata with copy-map in text

- a (black) source from the second column is copied to the corresponding (blue) target
- when a (black) source is missing the (blue) target is removed

```
scientificMetadata.instrument_name scientificMetadata.instrument.name.value
scientificMetadata.sample_name scientificMetadata.sample.name.value
scientificMetadata.instrument
scientificMetadata.sample
scientificMetadata.data
scientificMetadata.experiment_description
scientificMetadata.experiment_identifier
scientificMetadata.end_time scientificMetadata.end_time.value
scientificMetadata.start_time scientificMetadata.start_time.value
scientificMetadata.title scientificMetadata.title.value
scientificMetadata.program_name
```

Grouping scientificMetadata with group-map

The scan scientificMetadata can be grouped in measurement scientificMetadata by group-map

- a file containing the group-map is defined by **metadata_group_map_file** *(str)* , default: `None`
  `/home/p08user/.config/DESY/scingestor-metadata-group-map.lst`
- the group-map file contain `text` file with two or three column array,
  i.e. `<target> <source> [<target type>]`
- target type is not mandatory, and it can be:
  `Min`, `Max`, `MinMax`, `Range`, `List`, `UniqueList`, `Dict`, `First`, `Last`, `FirstLast`, `Endpoints`
- if the target type is missing for the numerical source values we store: `average`, `min`, `max`, `std`, `counts`
- if the target type is missing for the string source values unique string is stored

Group-map file structure

the (black) source from the second column is grouped to the corresponding (blue) target with use of the (green) target type

```
scientificMetadata.ScanCommand scientificMetadata.ScanCommand List
scientificMetadata.source_energy scientificMetadata.instrument.source.energy Dict
scientificMetadata.source_current scientificMetadata.instrument.source.current Range
scientificMetadata.sample_chemical_formula scientificMetadata.sample.chemical_formula.value
scientificMetadata.sample_temperature scientificMetadata.sample.temperature.value
scientificMetadata.sample_name scientificMetadata.sample.name.value
scientificMetadata.HKLmatrix scientificMetadata.user_data.flathklmatrix UniqueList
```

Measurement in Raw/Derived Datasets

The group measurement `Datasets` are stored in

- `DerivedDatasets`
  in case we store both scan Datasets and grouped (measurement) Datasets
- `RawDatasets`
  in case we don't store scan Datasets but only grouped (measurement) Datasets

The choice is made in the scingestor configuration by setting, e.g.

```
skip_scan_dataset_ingestion:  True
raw_metadata_callback:  True
```

Start Sardana scan: metadata only from NeXus

To start Sardana scan on `haspp08lisa2.desy.de` (P08 beamline)

- select NeXus configuration components e.g. in

```
nxselector
```

- run

```
spock
```

- in the spock command-line interface execute

```
scdd/door/hasvmscdd01.01 [1]:  senv NXSAppendSciCatDataset True
scdd/door/hasvmscdd01.01 [1]:  senv SciCatDatasetListFileLocal True
scdd/door/hasvmscdd01.01 [1]:  qscan 0.0 0.0 0.0 0.0 0.068 0.135 14 1.0
```

Also `ScanFile`, `ScanDir` and `ScanID` need to be defined

Support for FIO files and general hooks

- support for fetching Metadata from NeXus and FIO files
- default order of searching for master metadata file: **"nxs", "h5", "ndf", "nx", "fio"**

```
from sardana.macroserver.macro import Macro
from nxstools.pyeval import scdataset

class gh_post_scan(Macro):

   def run(self):
      scdataset.append_scicat_dataset(self)
```

Start Sardana scan: metadata from NeXus, FIO

To start Sardana scan on `haspp08lisa2.desy.de` (P08 beamline)

- select NeXus configuration components e.g. in

```
nxselector
```

- run

```
spock
```

- in the spock command-line interface execute

```
scdd/door/hasvmscdd01.01 [1]:  senv AppendSciCatDataset True
scdd/door/hasvmscdd01.01 [1]:  senv SciCatDatasetListFileLocal True
scdd/door/hasvmscdd01.01 [1]:  qscan 0.0 0.0 0.0 0.0 0.068 0.135 14 1.0
```

Also `ScanFile`, `ScanDir` and `ScanID` need to be defined

Measurement sardana macros

The measurement can be started/stopped with sardana macros manually

- `start_measurement <dataset_name>`
  starts a new measurment with the given name
- `make_measurement <dataset_name>`
  starts a new measurment with the given name and adds to the measurement the last scan
- `stop_measurement`
  updates the current measurement dataset in the SciCat database and stops the measurement
- `update_measurement`
  updates the current measurement dataset in the SciCat database
- `show_current_measurement`
  shows the current measurement name

AutoGrouping measurement mode

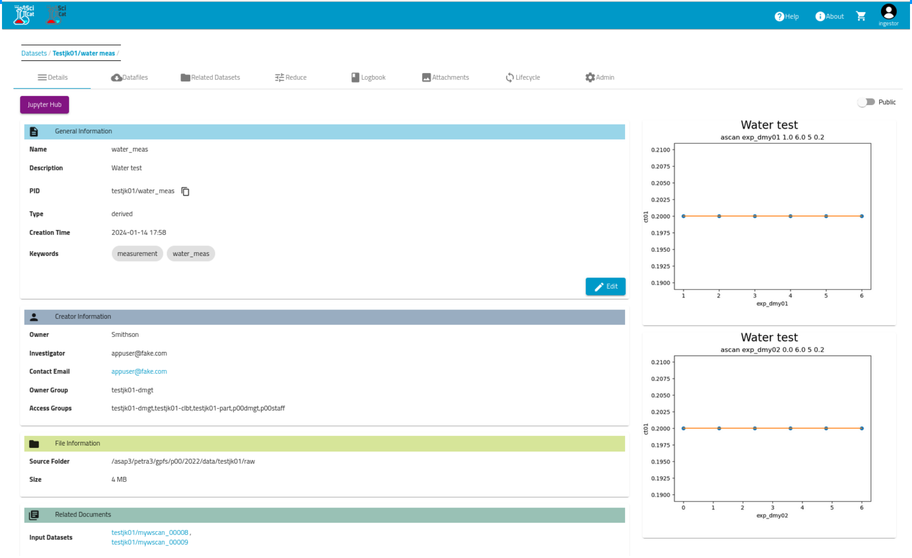
We can switch on the AutoGrouping mode by

```
scdd/door/hasvmscdd01.01 [1]:  senv SciCatAutoGrouping True
```

In this mode

- scan metadata is grouped automatically in the measurement dataset
- the measurement dataset updated after each scan
- the name of measurement is taken from the base scanname after removing ScanID,
  e.g. for `<scanname> = "mycalib2_00012"`
  the measurement name is `"mycalib2"`

Datasets in the SciCat frontend



Datasets / **Testjk01/water meas** /

General Information

Name	water_meas
Description	Water test
PID	testjk01/water_meas
Type	derived
Creation Time	2024-01-14 17:58
Keywords	measurement water_meas

Creator Information

Owner	Smithson
Investigator	appuser@fake.com
Contact Email	appuser@fake.com
Owner Group	testjk01-dmgt
Access Groups	testjk01-dmgt,testjk01-clbt,testjk01-part,p00dmgt,p00staff

File Information

Source Folder	/asap3/petra3/gpfs/p00/2022/data/testjk01/raw
Size	4 MB

Related Documents

Input Datasets	testjk01/mywscan_00008 , testjk01/mywscan_00009

Questions ?

Thank You !