2025-04-03 Meeting notes ESRF Ontologies

Date

03 Apr 2025

Attendees

- Wout De Nolf
- Marjolaine Bodin
- Pierre-Olivier Autran

Goals

- Status of the PANET/ESRF techniques - ontology and the plans for filling the information into icat?

Discussion items

Python project to use in Bliss and ICAT+ backend

https://esrf-ontologies.readthedocs.io

```
pip install esrf-ontologies
```

Bliss ICAT fields

```
from esrf_ontologies import technique

technique.get_technique_metadata("XRF", "XAS").get_dataset_metadata()
{'technique_pid': 'http://purl.org/pan-science/ESRFET#XAS http://purl.org/pan-science/ESRFET#XRF',
'definition': 'XAS XRF'}
```

Bliss HDF5/NeXus

```
from esrf_ontologies import technique

technique.get_technique_metadata("XRF", "XAS").get_scan_info()
{'techniques': {'@NX_class': 'NXnote', 'names': ['XAS', 'XRF'], 'iris': ['http://purl.org/pan-science
/ESRFET#XAS', 'http://purl.org/pan-science/ESRFET#XRF']}, 'scan_meta_categories': ['techniques']}
```

For the ICAT+ backend

```
from esrf_ontologies import technique
technique.get_all_techniques()[0]
Technique(iri='http://purl.org/pan-science/ESRFET#3DXRD', names=('3DXRD', '3D X-Ray Diffraction'),
description='3DXRD is a far field diffraction technique capable of reconstructing polycrystalline materials
from the individual diffraction spots recorded in the diffraction pattern given that the crystallographic
parameters of the material are known. The success of the technique is based on the ability of the individual
diffraction spots to be separated. This means there is a trade off between the number of grains and the degree
of strain that is measurable.')
```

Beamlines to implement it first

ID26

Nothing yet but Marius Retegan thinks we can easily add it for this beamline since techniques have their specific commands.

Code: https://gitlab.esrf.fr/bcu-vercors/ID26/id26

For example: https://gitlab.esrf.fr/bcu-vercors/ID26/id26/-/blob/master/id26/controllers/xas.py?ref_type=heads

ID32

Technique (RIXS, XES, XAS) is currently in the dataset name:

Example: https://data.esrf.fr/investigation/2111528864/datasets

Code: https://gitlab.esrf.fr/id32/id32 (I do not see any case the creates dataset names)

ID24

Technique field (XAS) is send to ICAT by the Bliss scripts:

Example: https://data.esrf.fr/investigation/2116235361/datasets

Code: https://gitlab.esrf.fr/ID24/id24 uncommitted changes

```
blissadm@foucault:~/local/id24.git$ grep -r definition\" .
./id24/HPLFsequence.py:        self._description["definition"] = SimpleSetting("hplf_seq_run_definition",
default_value="")
./id24/HPLFsequence.py:        #self._run_definition = SimpleSetting("hplf_seq_run_definition",
default_value="")
./id24/HPLFsequence.py:                "definition": self._description["definition"].get()
./id24/HPLFsequence.py:            res = str(self._description["definition"].get())
./id24/HPLFsequence.py:                if key == "definition":
./id24/HPLFsequence.py:            res = GREEN(self._description["definition"].get())
./id24/HPLFsequence.py:            val = self._description["definition"]
./id24/HPLFsequence.py:                elif key == "definition":
./id24/HPLFsequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/HPLFsequence.py:        setup_globals.SCAN_SAVING.dataset["Sample_name"] = self._description
["definition"].get()
./id24/HPLFsequence.py:        # setup_globals.SCAN_SAVING.dataset["definition"] = "BK50/CuZrAl23/Dia20.7 delay
(ns):-9 energy(J):10"    --> column 3 in the ICAT table
./id24/HPLFsequence.py:        # setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"         --> column 3
in the ICAT table
./id24/HPLFsequence.py:        # setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "UNKNOWN"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XMCD"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "UNKNOWN"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "UNKNOWN"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "UNKNOWN"
./id24/ID24sequence.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:            setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24sequence.py:            setup_globals.SCAN_SAVING.dataset["definition"] = "UNKNOWN"
./id24/ID24escan.py:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
./id24/ID24escan.py.before_fastscan:        setup_globals.SCAN_SAVING.dataset["definition"] = "XAS"
```

ID21

Technique field (XANES, XRF_2Dmap, ...) is send to ICAT by daiquiri:

Example: https://data.esrf.fr/investigation/2024055611/datasets?filters=

Code: https://gitlab.esrf.fr/search?search=additional_metadata&nav_source=navbar&project_id=4393&group_id=1073&search_code=true

Roadmap

milestone 1: replace PANET with ESRFET in ICAT db + ingester since the pid is now the full IRI

milestone 2 (end 2025): replace all non-official "definition" field values TOMO, XRF_2Dmap, ... (Check what beamlines are pushing these)

milestone 3 (end 2026): pyicat-plus will validate the "definition" / "technique_pid" fields and not allow monkey business

milestone 4: reduce the UNKNOWN fraction below x %

communicate the deadlines with the data working group. Something else? BCU meeting? Andrew Goetz Vicente Rey-Bakaikoa Jens Meyer

Decisions

Generic pid's: XRAYS and EM will be created by ingester when no pid and definition="UNKNOWN".

For now when you have definition!="UNKNOWN" and no pid: wait for milestone 3 and cleanup db later.

Processed data: when technique pid is not provided, take it from the raw dataset

Action items