FAIR Data Infrastructure for Condensed-Matter and Chemical Physics

FAIRmat

APPLICATION DEFINITIONS

May 16th 2022

Andrea Albino

Application Definition

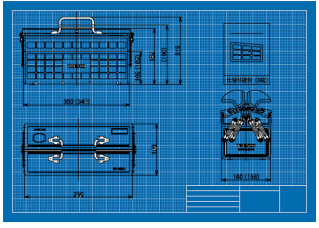
The schema

Covering Multiple Fields

Schema and Template concepts

SCHEMA

A formal description of data, data types, and data file structures, such as XML files.



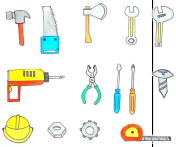
The blueprint of a toolbox

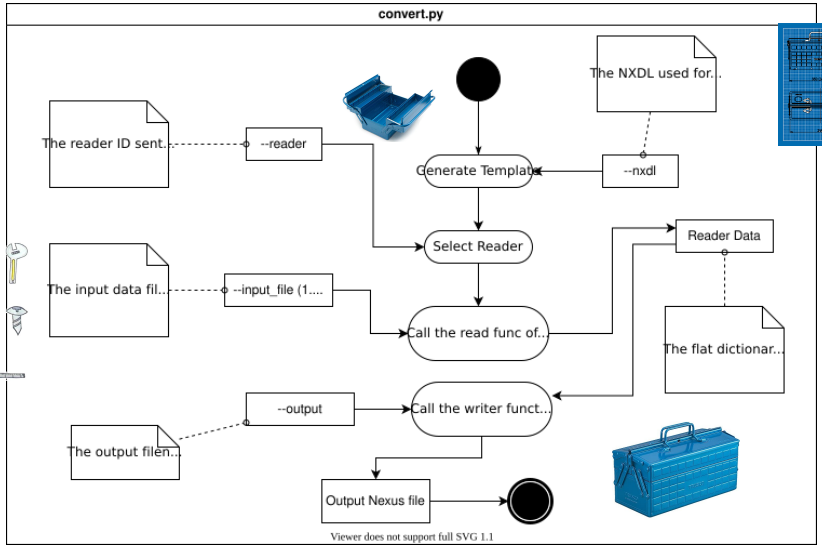
TEMPLATE

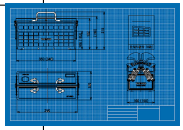
A physical object from which other objects are based or derived.

A toolbox that can be filled with a set of tools fitting in it

Area B Tools




convert.py
The reader ID sent...
--reader
The NXDL used for...
Generate Template
--nxdl
Select Reader
Reader Data
The input data fil...
--input_file (1....
Call the read func of...
The flat dictionar...
--output
Call the writer funct...
The output filen...
Output Nexus file
Viewer does not support full SVG 1.1



Area C, Area B and Area A

We can use pretty much the same tools.
Area B refers to communities using standard schemas (NeXus)
Area A needs to define some "standard" schema for Synthesis.
Area C could contaminate and enrich the Sample standard schema.

A standard enables and enhances exchange!

Even more: standard classes can be made *searchable* in NOMAD

The SAMPLE base class

symbols:

i: "Number of components in the first defined material"

j: "Number of components in the second defined material"

...

...

NXsample:

sample_id(NXid):

qualification:

set of measured properties...

NXsample_component:

exists: [min, 1, max, unbounded]

component_id(NXid):

qualification:

set of measured properties...

NXmaterial:

material_id(NXid):

components(NXid):

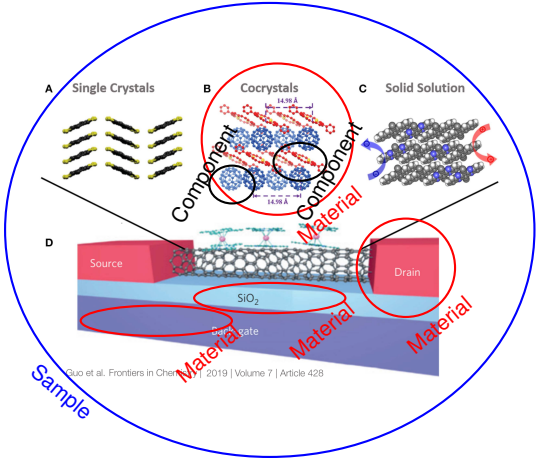
dimensions:

rank: 1

dim: [[1, i]]

qualification

set of measured properties... :



The SAMPLE base class

```
symbols:
    i: "Number of components in the first defined material"
    j: "Number of components in the second defined material"
    ...
    ...
        NXsample:
            sample_id(NXid):
            qualification:
            set of measured properties...
            NXsample_component:
                exists: [min, 1, max, unbounded]
                component_id(NXid):
                qualification:
                set of measured properties...
            NXmaterial:
                material_id(NXid):
                components(NXid):
                    dimensions:
                        rank: 1
                        dim: [[1, i]]
                qualification
                set of measured properties... :
```

a list of "components" placed outside "material"

components' ID are then linked in this list

The SAMPLE base class

symbols:

i: "Number of components in the first defined material"

j: "Number of components in the second defined material"

...

...

NXsample:

sample_id(NXid):

qualification:

set of measured properties...

NXsample_component:

exists: [min, 1, max, unbounded]

component_id(NXid):

qualification:

set of measured properties...

NXmaterial:

material_id(NXid):

components(NXid):

dimensions:

rank: 1

dim: [[1, i]]

qualification

set of measured properties... :

repeated fields into different classes

- ID
- qualification
- properties

may contain a "sample_component" or again a "material" (some inheritance is anyway likely to be adopted)

The SAMPLE base class

NXsample:

exists: [min, 1, max, unbounded]

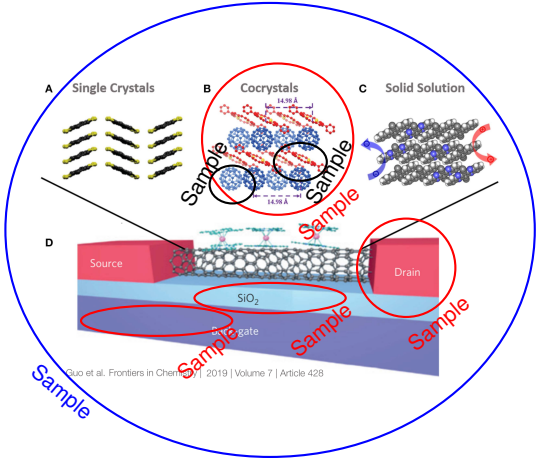
sample_id(NXid):

method: [experimental, simulation, declared by vendor]

qualification:

set of measured properties...

COMPONENT(NXsample):



The SAMPLE base class

NXsample:

exists: [min, 1, max, unbounded]

sample_id(NXid):

method: [experimental, simulation, declared by vendor]

qualification:

set of measured properties...

COMPONENT(NXsample):

- The term "sample" needs to be qualified:
- Simple Substance (Element or Chemical Compound)
- Mixture (or Single Phase Mixture)
  - (Solution, Suspension, Alloy or Colloid)
- Multi Phase Sample (Multi Domain or Composite)
- Layer (or Film or Substrate)
- Bulk
- Gel
- Dispersion
- Policrystalline Powder

The SAMPLE base class

NXsample:

exists: [min, 1, max, unbounded]

sample_id(NXid):

method: [experimental, simulation, declared by vendor]

qualification:

set of measured properties...

COMPONENT(NXsample):

- The set of fields is yet to be defined for each kind of sample

The SAMPLE base class

NXsample:

exists: [min, 1, max, unbounded]

sample_id(NXid):

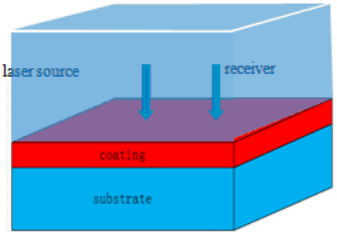
method: [experimental, simulation, declared by vendor]

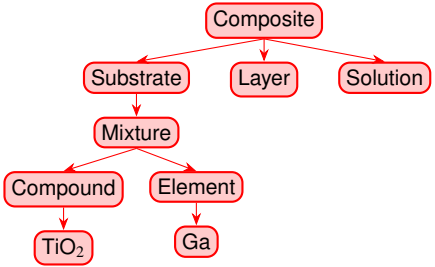
qualification:

set of measured properties...

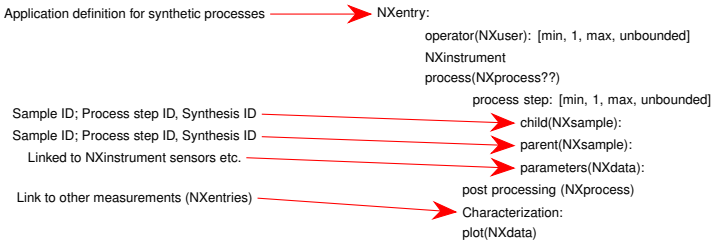
COMPONENT(NXsample):

- The sample tree hierarchy may be visualized in some way inside ELNs to have the whole composition of the sample at a glance





The SYNTHESIS base class



How to inherit these entities?

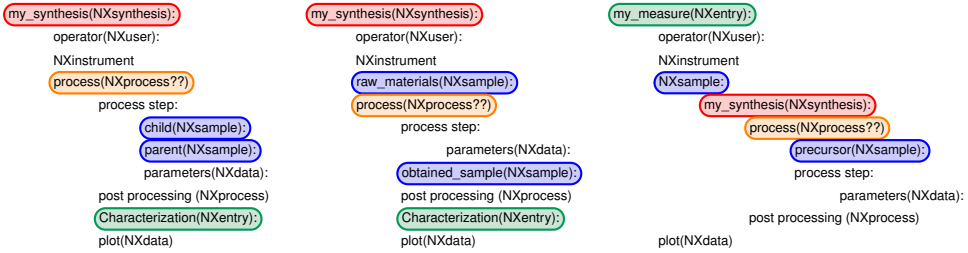
Sample

Synthesis

Measurement

Process

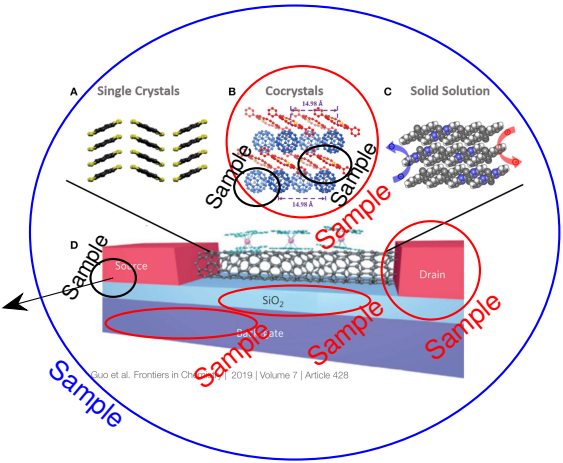
Multiple structures can be envisioned; each of them is plausible:



It would be enough to use a standard base class for SAMPLE and a custom one for SYNTHESIS

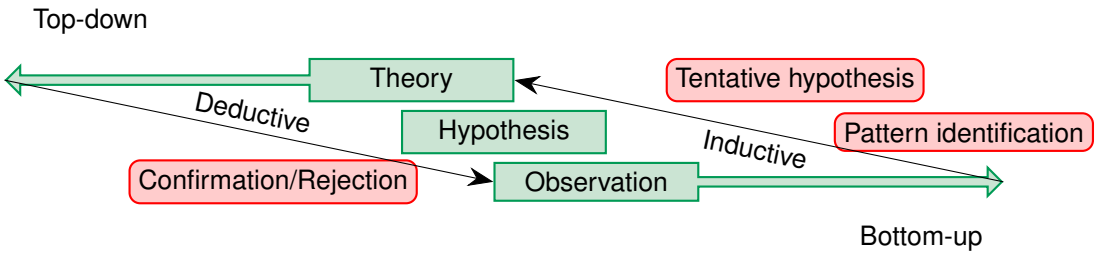
The SAMPLE base class

How to deal with interfaces ??
a new NXsample object could be instanciated, containing two parent samples



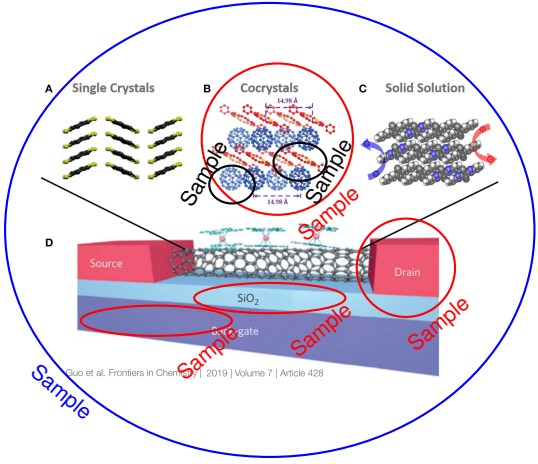
The SAMPLE base class

Collecting all these use cases and looking at their schemas could lead us to generalize it!



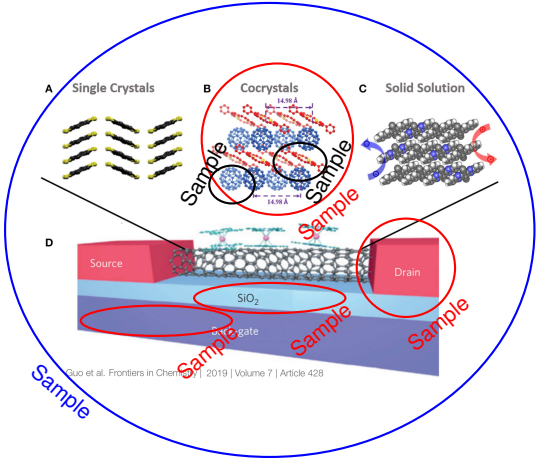
The SAMPLE base class

```
symbols:
    i: "Number of components"
    j: "Number of components"
    ...
    ...
        NXsample:
            exists: [min, 1, max, unbounded]
            sample_id(NXid):
            components(NXid):
                dimensions:
                    rank: 1
                    dim: [[1, i]]
            qualification:
            set of measured properties...
            COMPONENT(NXsample):
```



The SAMPLE base class

```
symbols:
    i: "Number of components"
    j: "Number of components"
    ...
    NXsample:
        exists: [min, 1, max, unbounded]
        sample_id(NXid):
        components(NXid):
            dimensions:
                rank: 1
                dim: [[1, i]]
        qualification:
        set of measured properties...
        COMPONENT(NXsample):
```



- We can avoid to list components as they come naturally from inherited structure
- We do not need to fix the terminology of multiscale components of our sample (this may add complexity and deprive of flexibility your sample definition)

The SAMPLE base class

NXsample:

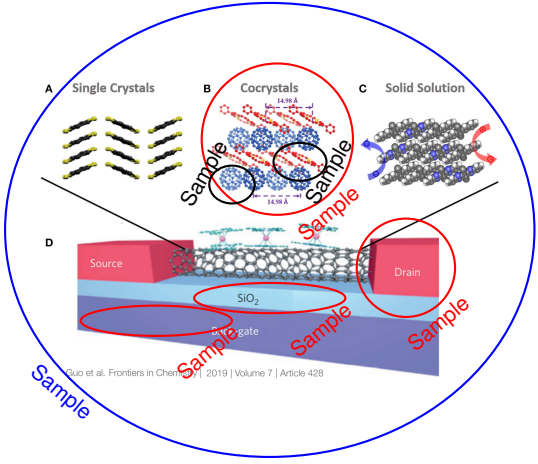
exists: [min, 1, max, unbounded]

sample_id(NXid):

qualification:

set of measured properties...

COMPONENT(NXsample):



The SAMPLE base class

NXsample:

exists: [min, 1, max, unbounded]

sample_id(NXid):

qualification:

set of measured properties...

COMPONENT(NXsample):

- The set of fields is yet to be defined
- The sample tree hierarchy may be visualized in some way inside ELNs to have the whole composition of the sample at a glance