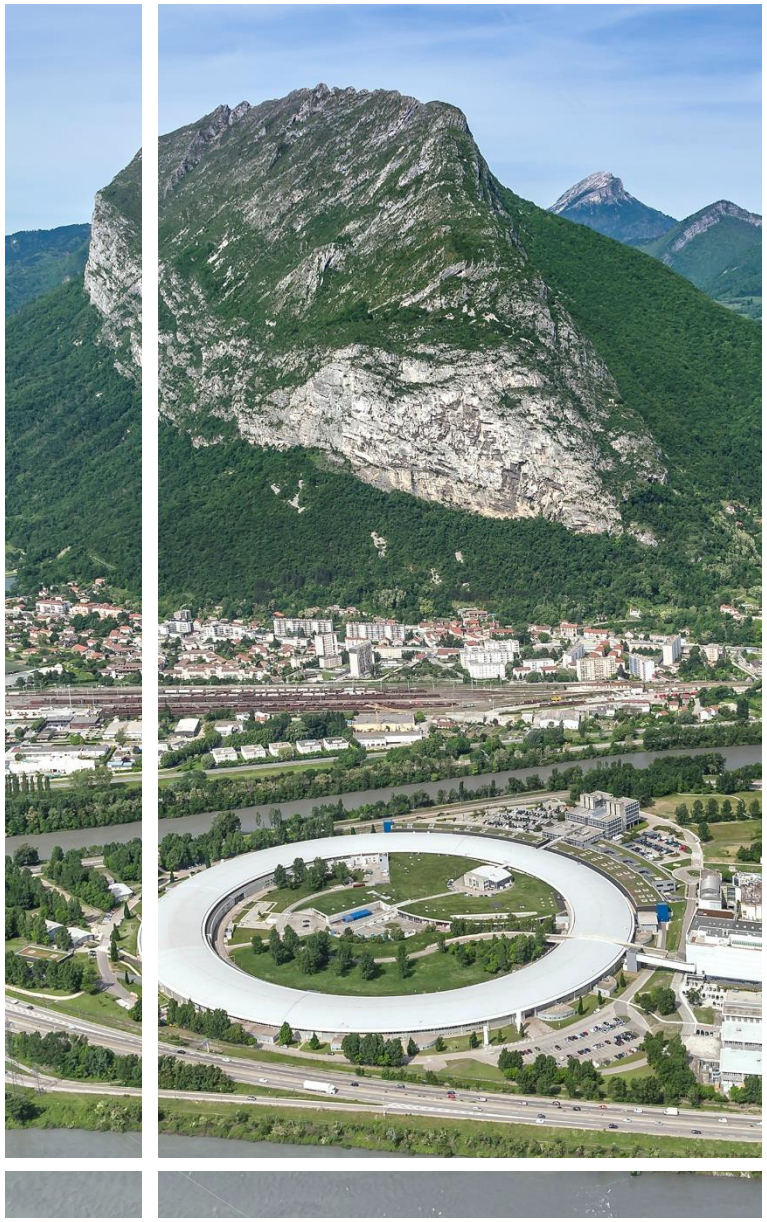

ESRF
The European Synchrotron



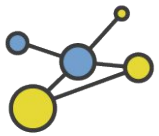
Ontological definition of experimental techniques for FAIR data

Wout De Nolf
ESRF (Data Automation Unit)

FAIR: provide meaning to data

Findable

Accessible



Interoperable

Reusable

Data Portal
https://data.esrf.fr

NeXus Data Format
https://www.nexusformat.org/

The infrastructure is in place but the “metadata” is missing.

FAIR: provide meaning to data

NeXus is a common data format for neutron, x-ray, and muon science.

nexus v2024.02 documentation » 3. NeXus: Reference Documentation » 3.3. NeXus Class Definitions » 3.3.2. Application Definitions previous | next | index

3.3.2. Application Definitions

A description of each NeXus application definition is given. NeXus application definitions define the *minimum* set of terms that *must* be used in an instance of that class. Application definitions also may define terms that are optional in the NeXus data file. The definition, in this case, reserves the exact term by declaring its spelling and description. Consider an application definition as a *contract* between a data provider (such as the beam line control system) and a data consumer (such as a data analysis program for a scientific technique) that describes the information is certain to be available in a data file.

Use NeXus links liberally in data files to reduce duplication of data. In application definitions involving raw data, write the raw data in the NXinstrument tree and then link to it from the location(s) defined in the relevant application definition.

NXarchive
This is a definition for data to be archived by ICAT (http://www.icatproject.org/).

NXarpes
This is an application definition for angular resolved photo electron spectroscopy.

NXcanSAS
Implementation of the canSAS standard to store reduced small-angle scattering data of any dimension.

NXdirecttof
This is a application definition for raw data from a direct geometry TOF spectrometer

NXfluo
This is an application definition for raw data from an X-ray fluorescence experiment

NXindirecttof
This is a application definition for raw data from a direct geometry TOF spectrometer

NXiqproc
Application definition for any $I(Q)$ data.

NXlauetof
This is the application definition for a TOF laue diffractometer

NXmonopd
Monochromatic Neutron and X-Ray Powder diffractometer

NXmx
functional application definition for macromolecular crystallography

NXrefscan
This is an application definition for a monochromatic scanning reflectometer.

NXreftof
This is an application definition for raw data from a TOF reflectometer.

NXsas
Raw, monochromatic 2-D SAS data with an area detector.

NXsastof
raw, 2-D SAS data with an area detector with a time-of-flight source

NXscan

Previous topic
3.3.1.58. NXxraylens

Next topic
3.3.2.1. NXarchive

This Page

Have a Question? Get help

Quick search
Go

https://www.nexusformat.org/
(since 2003)

3.3.2.11. NXrefscan

Status:

application definition, extends NXobject

Description:

▸ This is an application definition for a monochromatic scanning reflectometer. ...

Symbols:

The symbol(s) listed here will be used below to coordinate datasets with the same shape.

nP: Number of points

Groups cited:
NXdata, NXdetector, NXentry, NXinstrument, NXmonitor, NXmonochromator, NXsample, NXsource

Structure:

entry: (required) NXentry

title: (required) NX_CHAR ⇇

start_time: (required) NX_DATE_TIME ⇇

end_time: (required) NX_DATE_TIME ⇇

definition: (required) NX_CHAR ⇇

▸ Official NeXus NXDL schema to which this file conforms ...

instrument: (required) NXinstrument ⇇

SOURCE: (required) NXsource ⇇

type: (required) NX_CHAR ⇇

name: (required) NX_CHAR ⇇

probe: (required) NX_CHAR ⇇

Any of these values: `neutron` | `x-ray` | `electron`

monochromator: (required) NXmonochromator ⇇

wavelength: (required) NX_FLOAT {units=NX_WAVELENGTH} ⇇

DETECTOR: (required) NXdetector ⇇

data: (required) NX_INT (Rank: 1, Dimensions: [nP])

polar_angle: (required) NX_FLOAT (Rank: 1, Dimensions: [nP]) {units=NX_ANGLE}

sample: (required) NXsample ⇇

name: (required) NX_CHAR ⇇

Descriptive name of sample

rotation_angle: (required) NX_FLOAT (Rank: 1, Dimensions: [nP]) {units=NX_ANGLE} ⇇

3.3.2.12. NXreftof

Status:

application definition, extends NXobject

Description:

This is an application definition for raw data from a TOF reflectometer.

Symbols:

The symbol(s) listed here will be used below to coordinate datasets with the same shape.

xSize: xSize description

ySize: ySize description

nTOF: nTOF description

Groups cited:
NXdata, NXdetector, NXdisk_chopper, NXentry, NXinstrument, NXmonitor, NXsample

Structure:

entry: (required) NXentry

title: (required) NX_CHAR ⇇

start_time: (required) NX_DATE_TIME ⇇

end_time: (required) NX_DATE_TIME ⇇

definition: (required) NX_CHAR ⇇

▸ Official NeXus NXDL schema to which this file conforms ...

instrument: (required) NXinstrument ⇇

name: (required) NX_CHAR ⇇

chopper: (required) NXdisk_chopper

distance: (required) NX_FLOAT {units=NX_LENGTH} ⇇

Distance between chopper and sample

detector: (required) NXdetector ⇇

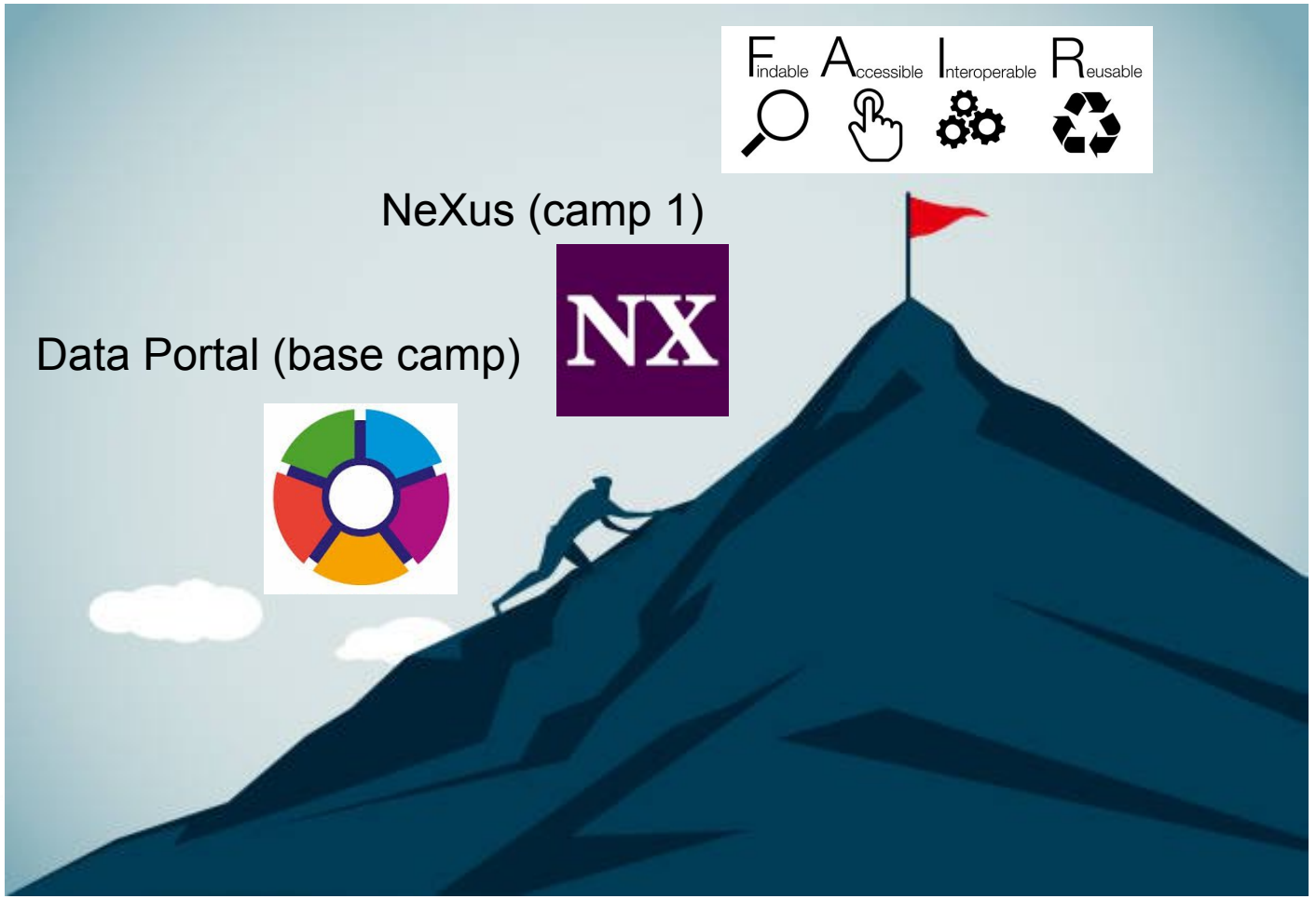
data: (required) NX_INT (Rank: 3, Dimensions: [xSize, ySize, nTOF])

time_of_flight: (required) NX_FLOAT (Rank: 1, Dimensions: [nTOF]) {units=NX_TIME_OF_FLIGHT} ⇇

▸ Array of time values for each bin in a time-of-flight ...

Provide meaning to data: climbing Mount FAIR

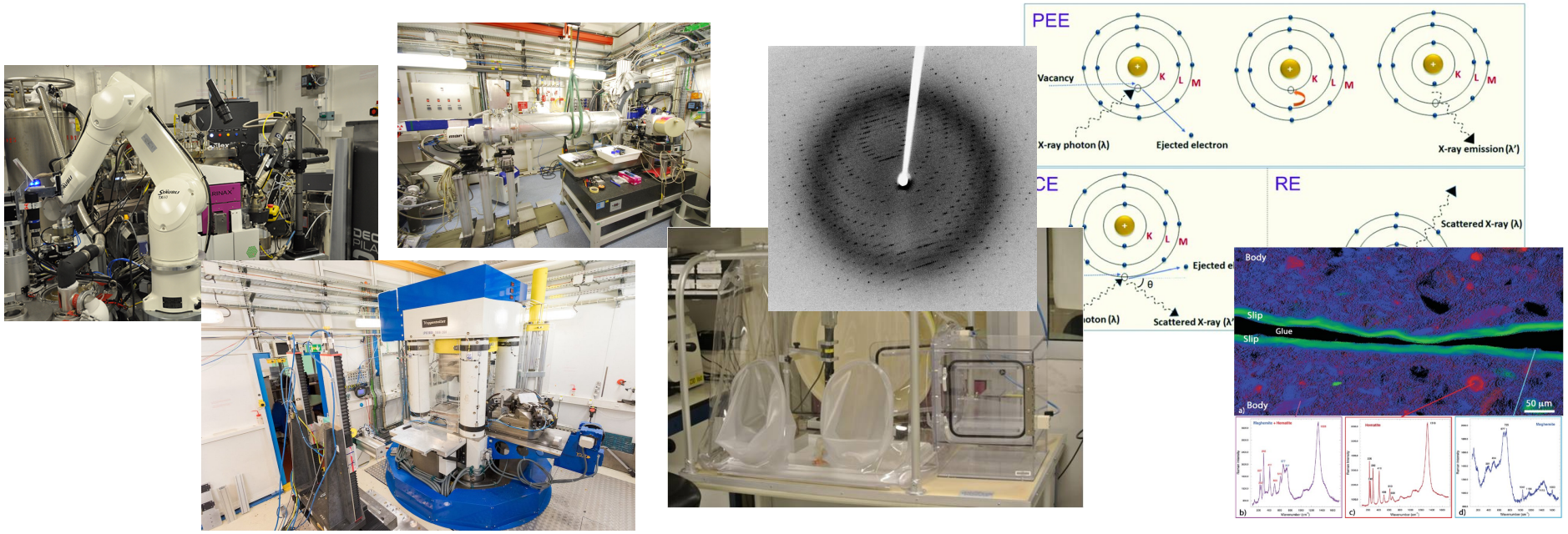
Why are we here?



Provide meaning to data: mountaineering 101

Why are we here?

Sense of **dread** when thinking about metadata covering instrumentation, sample preparation/origin, data acquisition, data processing, ...



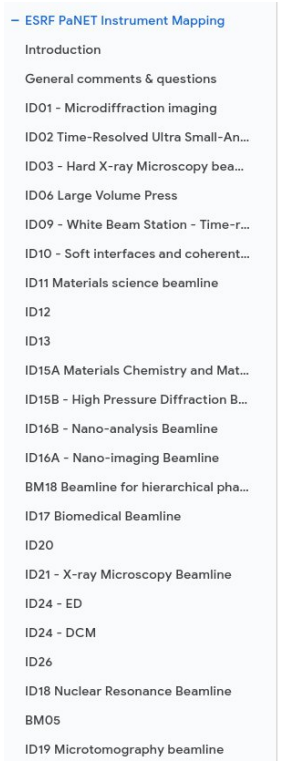
So how do you start climbing Mount FAIR?

Start by reducing the scope in which to define metadata.

In other words, start by **defining** techniques.

Provide meaning to data: experimental techniques

Option 1: Flat list of technique names per beamline and per facility



ESRF PaNET Instrument Mapping

ID24 - ED

Current names	PaNET	Missing techniques or remarks
EXAFS - extended X-ray absorption fine structure	EXAFS https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01198&jump_to_nav=true	
FTIR - Fourier transform infrared spectroscopy/microscopy	FTIR https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01320&jump_to_nav=true	
XANES - X-ray absorption near-edge structure	XANES https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01199&jump_to_nav=true	
XAS - X-ray absorption spectroscopy	XAS https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01196&jump_to_nav=true	
XMCD - X-ray magnetic circular dichroism	XMCD https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01137&jump_to_nav=true	

ID24 - DCM

Current names	PaNET	Missing techniques or remarks
EXAFS - extended X-ray absorption fine structure	EXAFS https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01198&jump_to_nav=true	
FTIR - Fourier transform infrared spectroscopy/microscopy	FTIR https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01320&jump_to_nav=true	
MicroXANES - micro X-ray absorption near-edge structure	XANES https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01199&jump_to_nav=true	missing micro ?
XAS - X-ray absorption spectroscopy	XAS https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01196&jump_to_nav=true	
XMCD - X-ray magnetic circular dichroism	XMCD https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=http%3A%2F%2Fpurl.org%2Fpan-science%2FPaNET%2FPaNET01137&jump_to_nav=true	

No meaning, just technique names

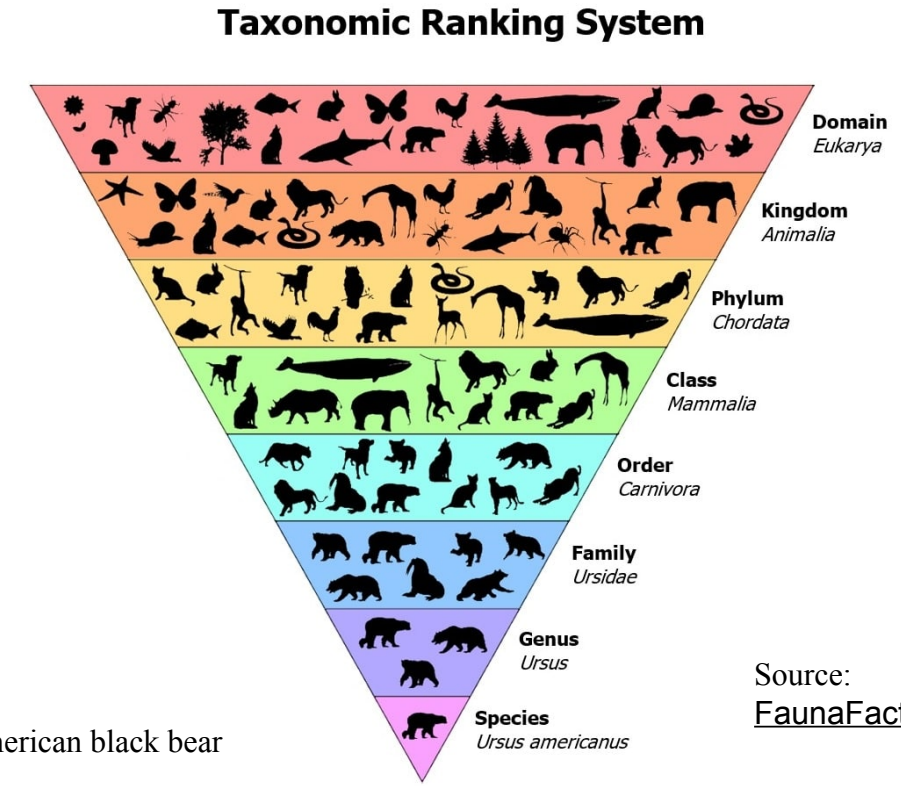
HR-XRPD at ESRF-ID22 may not be the same as at other beamlines.

You cannot relate techniques to each other and with other scientific fields.

Courtesy: Renaud Duyme (ESRF)

Option 2: **Taxonomy**, define techniques with a structured and hierarchical classification

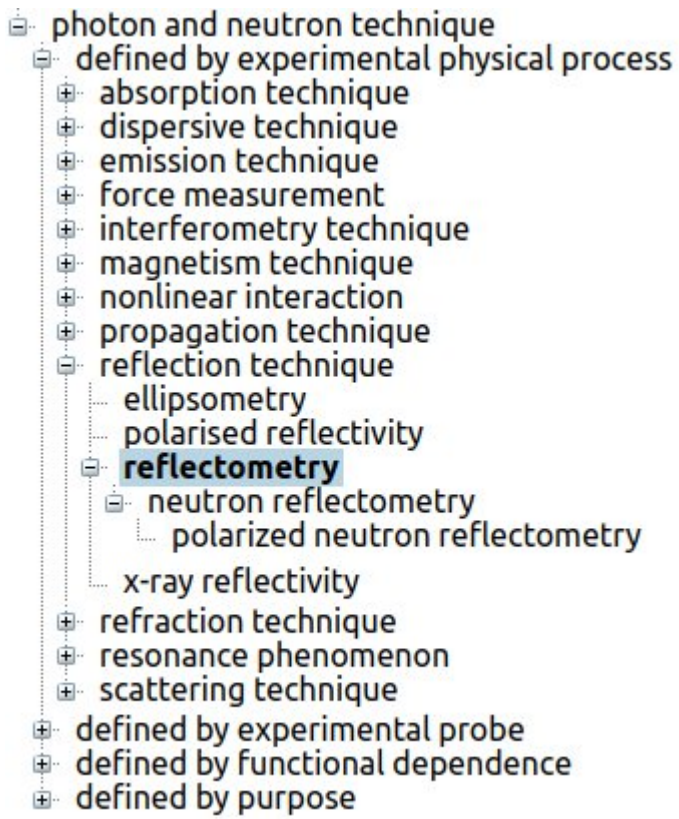
For example in Biology:
Linnaean Classification System



Source:
FaunaFacts

X-ray Powder diffraction ← American black bear

Provide meaning to data: context of the experimental techniques



Option 2: **Taxonomy**, define techniques with a structured and hierarchical classification

The PaNET ontology provides a taxonomy and thesaurus of photon and neutron (PaN) experimental techniques (developed in ExPaNDS context).

A tree of subclasses relates techniques without actually specifying what this relation is. In other words, it still does not contain enough meaning.

https://bioportal.bioontology.org/ontologies/PANET/?p=classes&conceptid=root

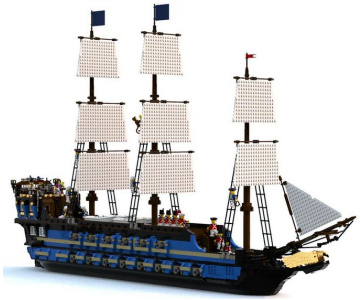
Provide meaning to data: ontology of experimental techniques

Option 3: **Ontology**, where we can:

- Use basic building blocks to **compose/define techniques** just like building castles, boats, cars etc. with LEGO blocks (Description-Logic).
- Relationships between techniques are **automatically inferred** (Reasoning based on Description-Logic) and can guide the creation of building blocks.
- Techniques need to be defined only to the extent that they can be distinguished from other techniques (**Differential meaning**).
- Utilize/connect to PaNET and other relevant ontologies in the domains of material science, physics, biology etc. (**relations provide meaning**).

Provide meaning to data: developing an ontology

Techniques



Building blocks

Knowledge engineer + scientist

Try to building your techniques with existing blocks.

Create new blocks when needed. *I want to build a truck but I don't have wheels.*

Techniques need to be **defined only to the extent that they can be distinguished** from other techniques. *The first person who makes a ship does not need to think about frigates, battleships, cogs, ...*

Provide meaning to data: developing an ontology

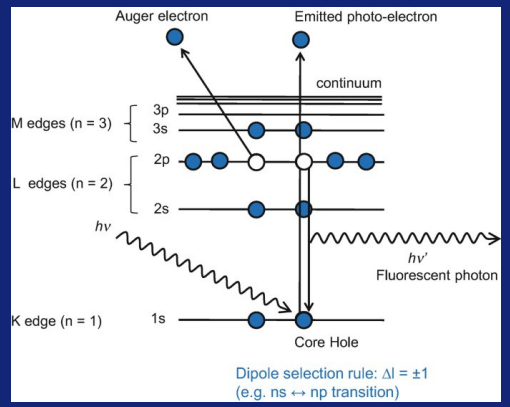
Laser-driven shock compression X-ray Absorption Spectroscopy

(for sake of illustration only)

X-rays

Laser

sample



X-rays

Electrons

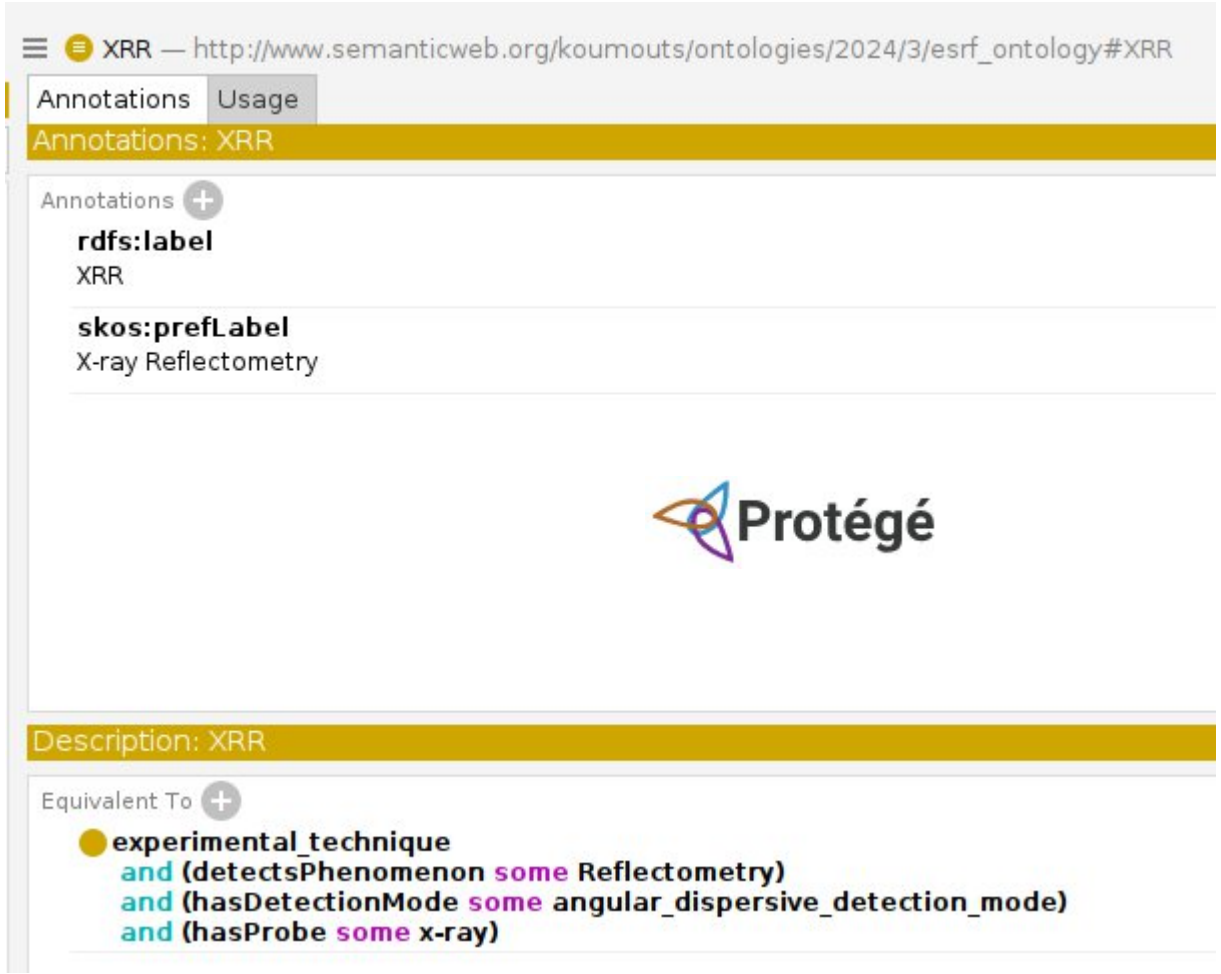
Detection:
Fluorescence, transmission, auger electrons, total electron yield, ...
High-resolution, energy-dispersive, ...

Sample input:
- **X-rays**
- **Lasers**

Process:
- X-ray **absorption**
- **Compression**

Space/Time:
- As a function of the energy (**spectroscopy**)
- Different projections (tomography)
- Pulsed (time-resolved)

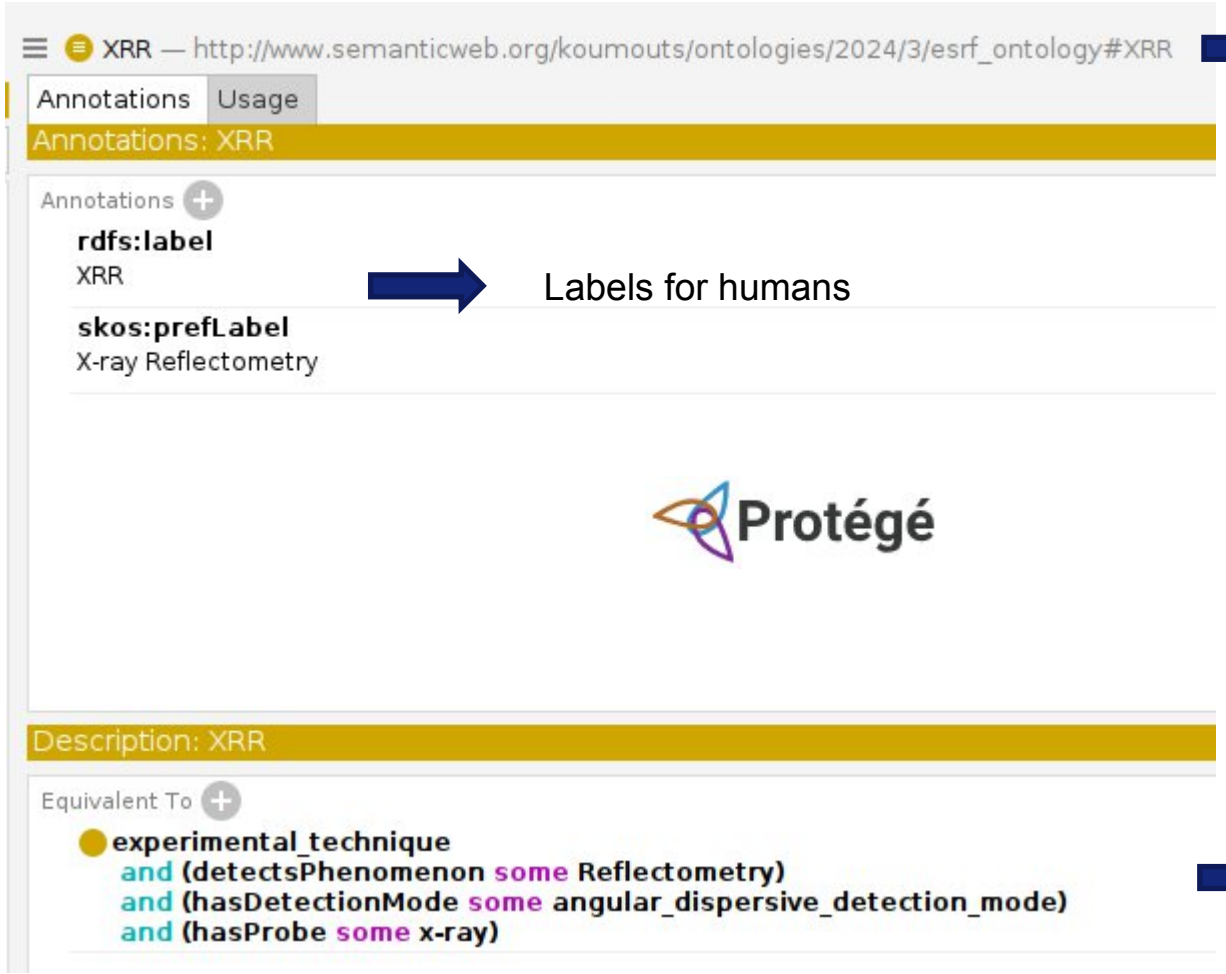
Provide meaning to data: practical implementation



Experimenting with Ontological building blocks (DL: classes, object properties, data properties)

Courtesy: Ioannis Koumoutsos (ESRF)

Provide meaning to data: practical implementation



IRI (Internationalized Resource Identifier)

→ persistent uniform resource locator

(https://purl.org/)

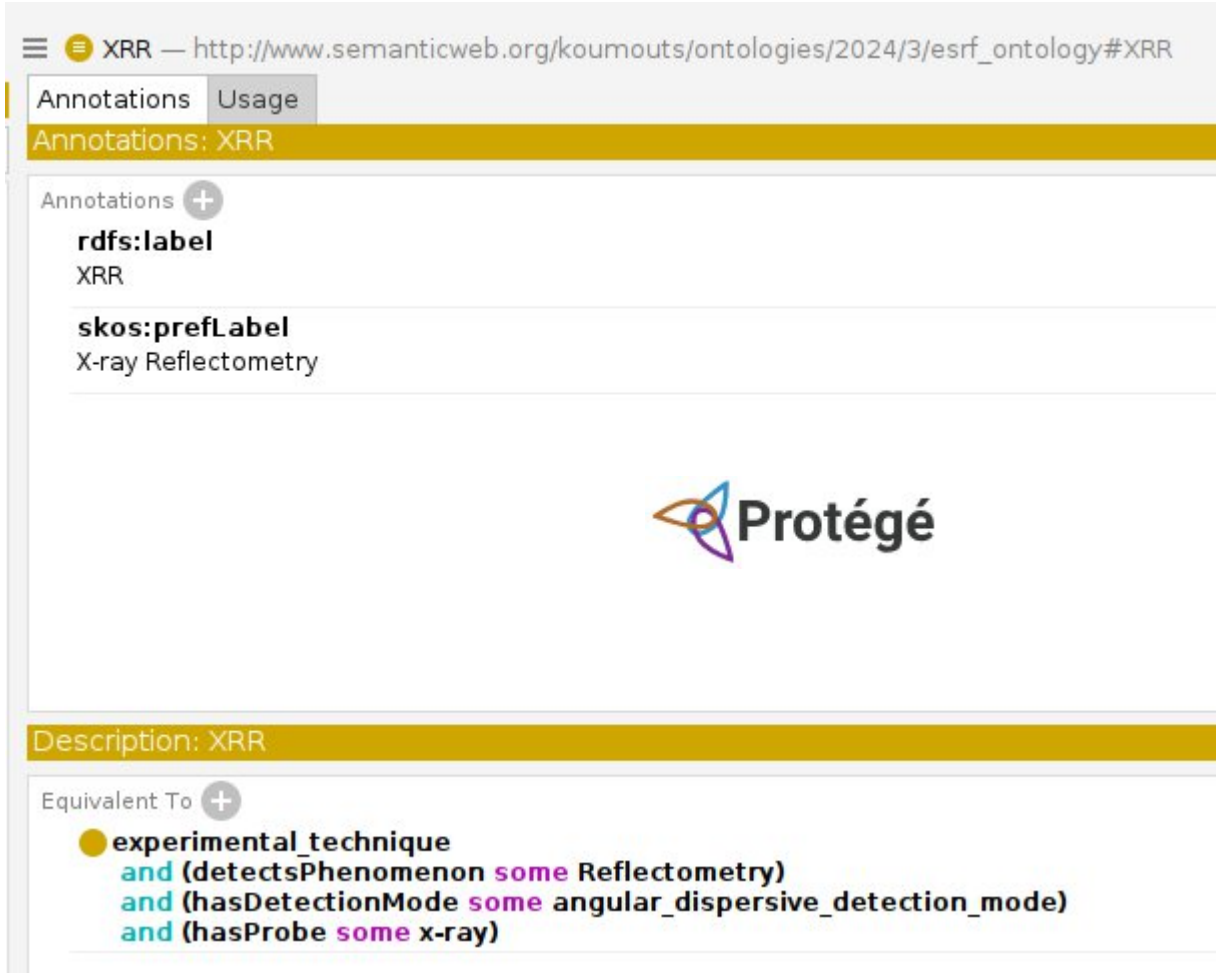
→ ⚠ Findable in FAIR ⚠

Labels for humans

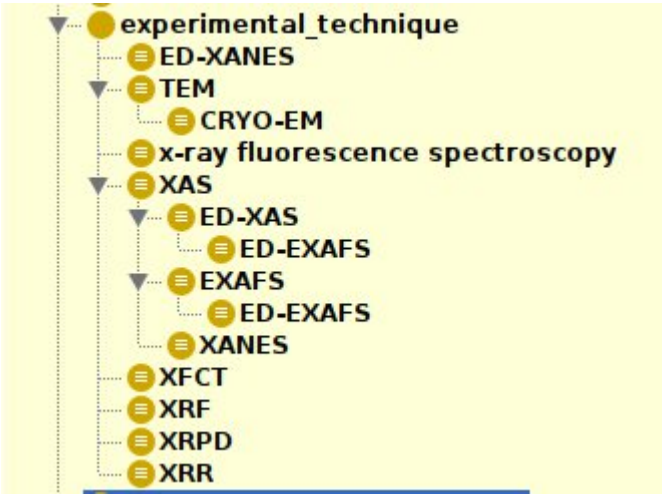
Building blocks (⚠ *just for illustration* ⚠)

Courtesy: Ioannis Koumoutsos (ESRF)

Provide meaning to data: practical implementation



Infer relationships by a reasoner based on the description logic

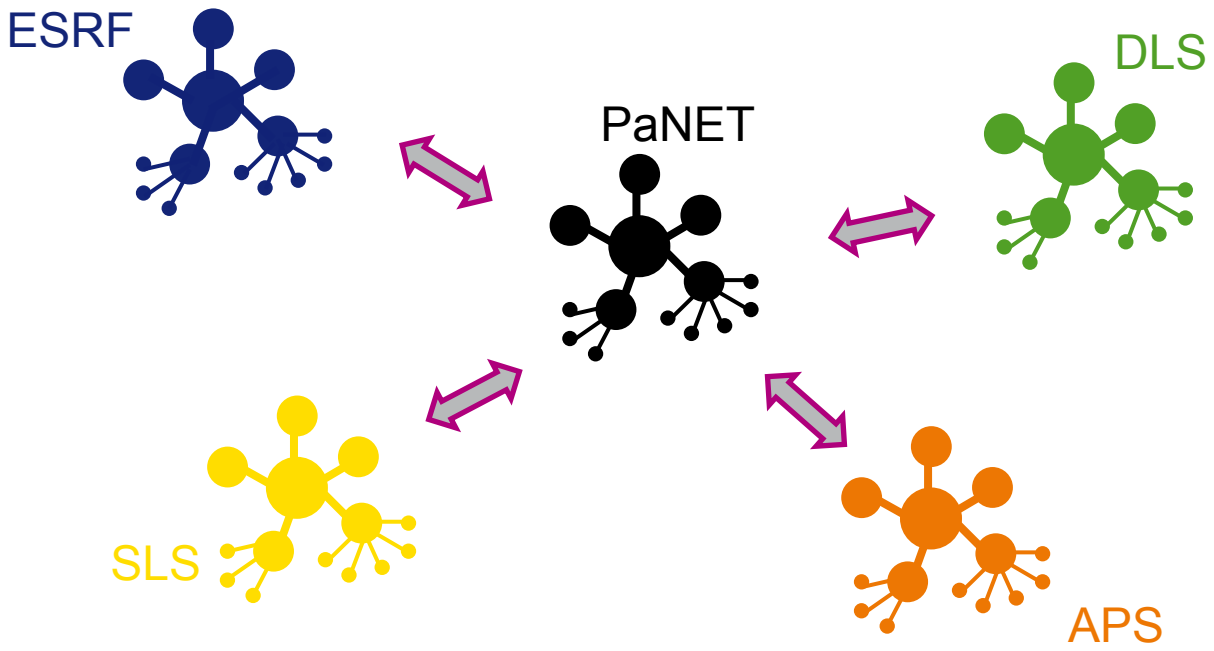


Courtesy: Ioannis Koumoutsos (ESRF)

Provide meaning to data: developing an ontology

Define a **common strategy/roadmap** on how to build and maintain ontologies at each institute and influence/connect to a central ontology (PaNET).

A **common vocabulary and meaning** arises from the connections.



NeXus is also defining an ontology

Provide meaning to data: developing an ontology

Short term goal at the ESRF:

Provide a *reasonable* way of defining techniques so all the data we publish is tagged with a technique IRI (in HDF5/NeXus and in ICAT).

- Adding new techniques in a matter of minutes without the need for endless discussions.
- Connect/influence the PaNET ontology over time.

✓Description Logic ✗Taxonomy

Long term goals:

- Add NeXus definitions for more techniques informed by existing connections between (meta)data and techniques.
- Improve or find inconsistencies in the current NeXus definitions.
- Infer techniques from (meta)data and vice versa.
- Connect to other ontologies to answer questions like
  - *I want to find out which pigments a 19th century painter used in this painting. Which technique can I use for that?*
  - *Where do I go to perform such an experiment?*
  - *Who do I contact to know whether this is the right technique for my particular use case?*
  - *Am I using the same technique as in this particular scientific publication?*
  - *How are X-ray powder diffraction and X-ray absorption spectroscopy related? What is the overlap and the differences in terms of characteristics and scientific problems they can solve?*
  - *Give me a list of non-existing techniques that could solve scientific problems that are not covered.*

✓Description Logic ✗Taxonomy