Photon and Neutron Ontologies from design to implementation

The study of what exists

Modern usage is the use of computer logic to represent knowledge in a certain domain

A cornerstone of the semantic web and modern knowledge representation

Photon and neutron PaN ontologies
dcterms:creator [type: xsd:string] A. Gonzalez Beltran
dcterms:creator [type: xsd:string] D Iyayi
dcterms:creator [type: xsd:string] H Gorzig orcid.org/0000-0001-9121-8643
dcterms:creator [type: xsd:string] S P Collins orcid.org/0000-0001-5120-0764
dcterms:creator [type: xsd:string] S. Da Graca Ramos

HZB Helmholtz Zentrum Berlin

diamond

1

ExPaNDS Task 3.2: Develop EU Photon and Neutron Ontologies

Lead partner: DLS, Contributors: ALBA, DESY, HZB, MaxIV, PSI, Soleil, SFTC

Develop ontologies *for main application domains of Photon and Neutron science to standardise the metadata used in* ***metadata catalogues*** *based on requirements defined in WP2. This will ensure that federated EOSC metadata catalogues are not only based on a common syntax, but also on a* ***common semantics****.*

For physical parameters we have NeXus.

Development of the ontology will be closely linked to the existing ***NeXus file format****…*

ExPaNDS European Open Science Cloud Photon and Neutron Data Services

panosc photon and neutron open science cloud

HZB Helmholtz Zentrum Berlin

diamond

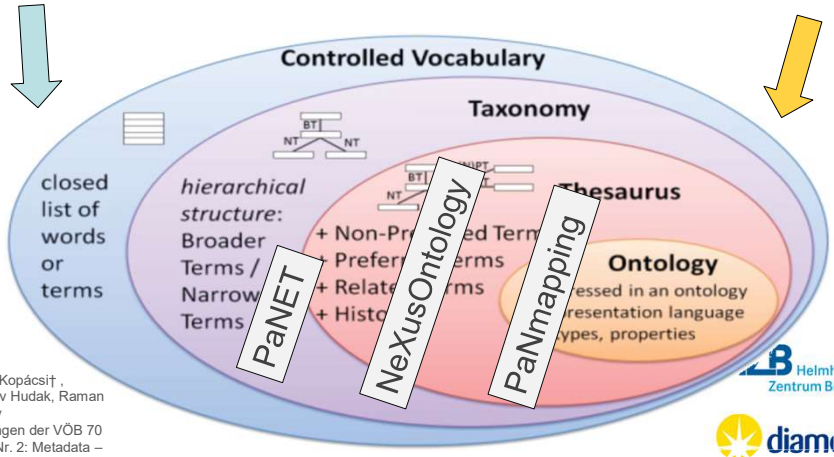
2

What is the right level of semantic richness?

Quick
Cheap
Easy
Local

Find a balance that is affordable, satisfies immediate use-cases and is extensible in the future

Slow
Expensive
Complex
Global
Interoperable



Sándor Kopácsi† , Rastislav Hudak, Raman Ganguly
Mitteilungen der VÖB 70 (2017) Nr. 2: Metadata – Metadaten

Helmholtz Zentrum Berlin

diamond

3

PaNET (Photon and Neutron Techniques Ontology)

HZB Helmholtz Zentrum Berlin

diamond

4

Purpose: PaNET (Photon and Neutron Techniques Ontology)

- An ontology to support **catalogue services** (i.e. Findability)
- Provide a controlled vocabulary of technique names with **Global Persistent Identifiers**
- Provide **alternate names** and annotation (e.g. Wikipedia entries)
- Provide class/**subclass** relationships
- Allow **semantic mapping** to similar or equivalent concepts in other vocabularies

Example: Tagging a dataset with 'MAD' should label it as 'multiwavelength anomalous diffraction'. It will then be found by searches including 'x-ray probe', 'macromolecular crystallography', 'single crystal diffraction', 'atomic core excitation' *etc*.

HZB Helmholtz Zentrum Berlin

diamond

5

Design Principles: PaNET (Photon and Neutron Techniques Ontology)

- Keep as simple as possible for current use-case while building in extensibility and community engagement
- Express as OWL
- Everything in the ontology is a PaN Technique
- Techniques are classes in the ontology
- Techniques are defined as being subclasses of multiple technique classes
- Start off with a small set of (almost) orthogonal basic high-level concepts and create trees of subclasses
- These correspond to 'implied' object properties that can be formalized later (but are not required for the immediate purpose).

photon and neutron technique
defined by experimental probe
photon probe
x-ray probe
defined by experimental physical process
defined by functional dependence
defined by purpose

HZB Helmholtz Zentrum Berlin

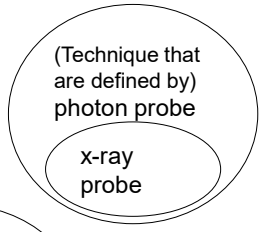
diamond

6

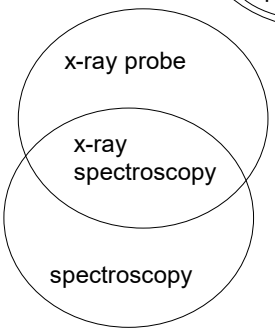
Design Principles: PaNET (Photon and Neutron Techniques Ontology)

Defining techniques by their (multiple) superclasses

- Refine by using subclasses (subsets):



- Refine by using multiple superclasses (set intersection):



HZB Helmholtz Zentrum Berlin

diamond

7

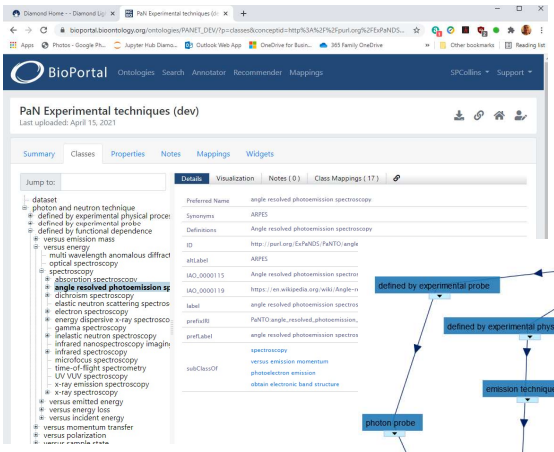
Implementation: PaNET (Photon and Neutron Techniques Ontology)

- Definitions entered on a **spreadsheet** (superclasses, name, alternate names, comment, Wikipedia reference etc)
- **ROBOT** tool creates an OWL ontology automatically
- Ontology reasoner creates **inferred subclass** relations
- Ontology uploaded to **NCBO Bioportal** ontology repository (an EOSC service), supporting SPARQL endpoint and REST API
- Ontology also accessible via **Github** (simple Python script will find information about a term)
- Ontology terms (i.e. techniques) directly resolvable via **WIDOCO** tool by typing their IRI into a web browser

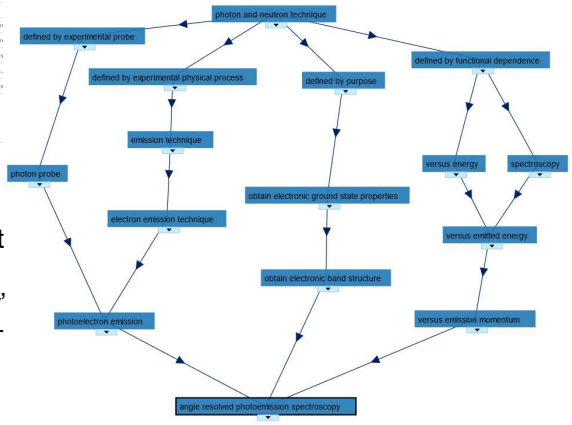
HZB Helmholtz Zentrum Berlin

diamond

8



Bioportal: Look up Angle Resolved Photoemission Spectroscopy

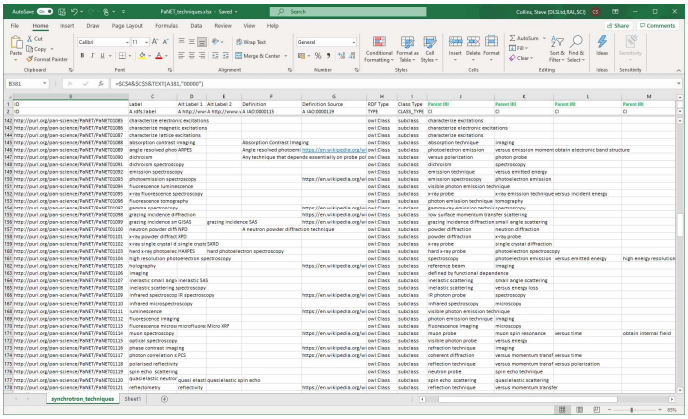


Catalogues search: A dataset tagged with 'Angle Resolved Photoemission Spectroscopy' will be found by searching for any of these 15 terms and their alternate names.

9

Community update and Development: PaNET (Photon and Neutron Techniques Ontology)

- Techniques are defined in a spreadsheet – very easy to share
- New techniques and modifications will be via a github workflow



- We invite the community to gradually add techniques, standard references etc to build a valuable resource for PaN science

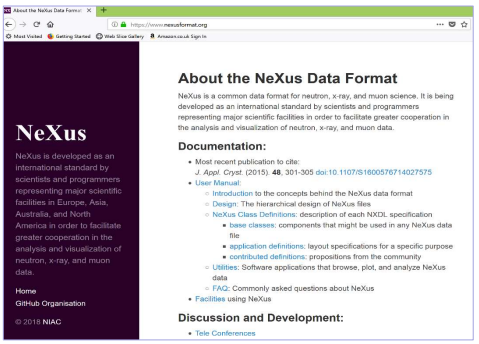
HZB Helmholtz Zentrum Berlin

diamond

10

And now, for something completely different...

NeXusOntology: An ontology of NeXus definitions



11

Purpose: NeXusOntology

- An ontology to support **catalogue services** (i.e. Findability)
- Flatten the NeXus class-by-class dictionaries into a single **controlled vocabulary**
- Provide NeXus terms with Global **Persistent Identifiers**
- A tool to visualize and **explore** the structure of NeXus classes, fields and attributes with NeXus annotations and links to NeXus documentation.
- Allow **mapping** to similar or equivalent concepts in other vocabularies

e.g. the NeXus field 'name' within the NeXus base class 'NXuser' is given a label 'NXuser-name' and a Global PID

`http://purl.org/nexusformat/definitions/NXuser-name`

(This is then mapped to `foaf:name` by our PaNmapping ontology)

HZB Helmholtz Zentrum Berlin

diamond

12

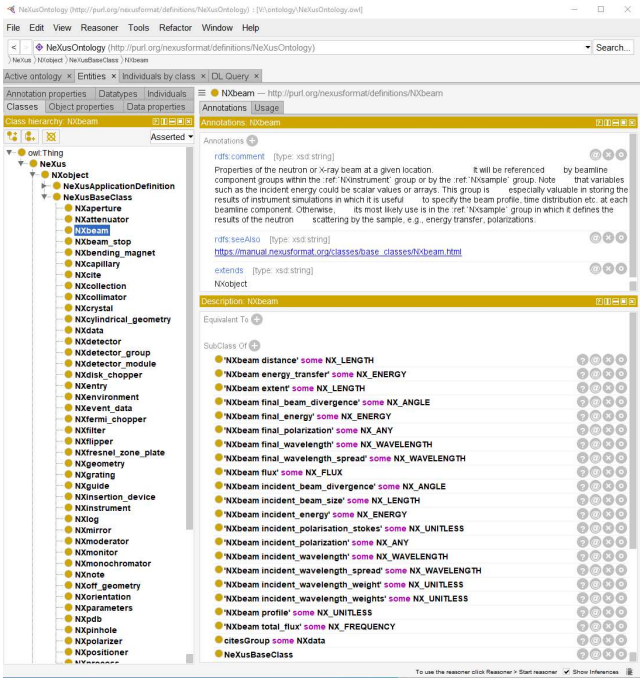
Design Principles: NeXusOntology

- NeXus ontology is an ontology of **NeXus definitions** (not an ontology of beamlines *etc*, or a representation of a NeXus file)
- The ontology formalizes NeXus definitions and (deliberately) does not introduce **new semantics**
- Because the only input to the ontology are the NeXus NXDL XML definition files, creation of the ontology is **automatic**
- We make explicit the **relationships** between Base Classes and fields etc
- **Persistent identifiers** are derived from NeXus names
- We expose NeXus **documentation** strings and provide links to individual fields in the NeXus on-line manual
- The ontology is essentially 'owned' by the **NeXus International Advisory Committee** (NIAC)

HZB Helmholtz Zentrum Berlin

diamond

13



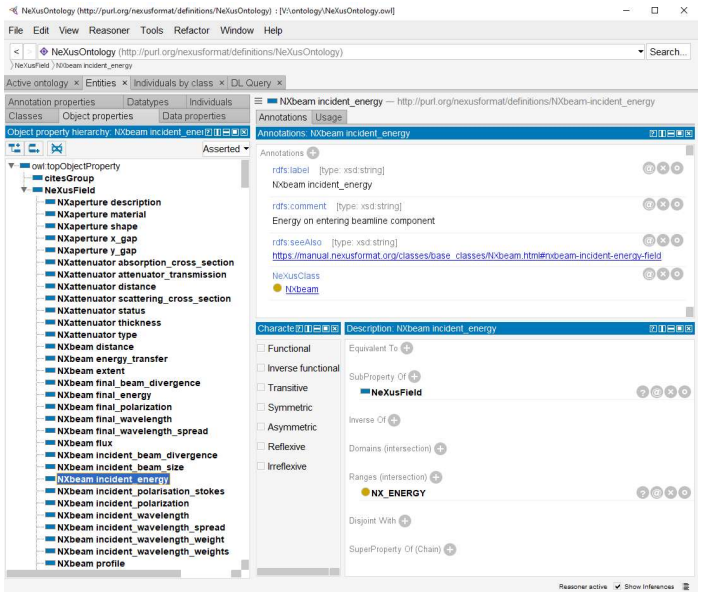
NeXusOntology viewed in Protégé: NeXus base classes (expressed as OWL classes)

HZB Helmholtz Zentrum Berlin

diamond

14

NeXusOntology viewed in Protégé: NeXus fields (expressed as OWL object properties)



15

Implementation: NeXusOntology

- A **Python script** parses the NXDL XML files and creates an OWL ontology using the owlready2 Python module
- A **purl.org** namespace is used for persistent identifiers
- A new version of the ontology is created each time the NeXus **NIAC** release a new version of NeXus definitions on the nexusformat github site
- The ontology can (in the future) be uploaded to an **ontology service** and documentation tools such as WIDOCO can allow machine-resolvable terms

16

Community update and Development: NeXusOntology

- Community maintenance is via the NeXus International Advisory Committee (**NIAC)**

17

PaN mapping v0.1

Aligning existing PaN-ontologies

Purpose

In the PaN-community different ontologies like the NeXus ontology, the CSMD of the ICAT and the PaNET have been created. All these ontologies are having there special usage scenarios nevertheless parts of the used concepts overlap. These concepts need to be related to each other.

In order to integrate these existing ontologies within the wider EOSC landscape, the existing ontologies need to be integrated with existing catalogue standards as e.g. DCAT v2 but also DublinCore, FOAF, PROV-O, and schema.org.

Caveats
This ontology is the very first version of integrating these ontologies. Contributors of the related ontologies might have other opinions of the relations and there might be changes in the future.

18

PaN mapping v0.1 - continued

Semantic tools

Simple Knowledge Organisation System (SKOS)

- skos:broader/skos:narrower for creating hierarchies
- skos:mappingRelation, skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch, or skos:relatedMatch

Web Ontology Language (OWL)
owl:equivalentClass

Resource Description Framework (RDF)
rdfs:subClassOf

HZB Helmholtz Zentrum Berlin

diamond

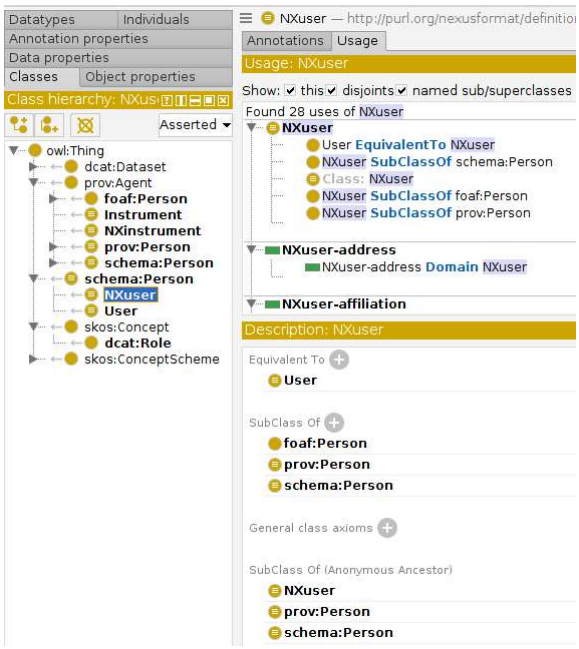
19

NXuser	equivalentTo, skos:exactMatch	csmd:User
	subClassOf	prov:Person, schema:Person, foaf:Person
NXuser:address	equivalentTo	schema:address
NXuser:affiliation	equivalentTo	schema:affiliation
NXuser:email	equivalentTo	schema:email, foaf:mbox
NXuser:fax_number	equivalentTo	schema:faxNumber

HZB Helmholtz Zentrum Berlin

diamond

20



21

Ten Simple Rules for Making a Vocabulary FAIR - evaluation

Rule 1. Determine the governance arrangements and custodian of the legacy vocabulary
- NeXus → NIAC will be custodian
- PaNET → DLS will be custodian
- PaN-mapping → HZB will be custodian

all: changes by community via github

Rule 2. Verify that the legacy-vocabulary license allows repurposing, and agree on the license for the FAIR vocabulary
- NeXus → GNU license
- PaNET → CC-BY 4.0 license
- PaN-mapping → CC-0 license

Rule 3. Check term and definition completeness and consistency in the legacy vocabulary
- NeXus → The term definitions refer to NeXus documentation and annotation.
- PaNET → All the terms have a label, and some have definitions.
- PaN-mapping → No new terms have been created.

Rule 4. Establish a technical maintenance environment for the FAIR vocabulary
- NeXus → NeXusOntology is auto-generated from NeXus NXDL definition files.
- PaNET → Ontology development and maintenance processes are using GitHub.
- PaN-mapping → Development and maintenance processes are using Protégé and GitHub.

Rule 5. Assign a unique and persistent identifier to (a) the vocabulary and (b) each term in the vocabulary
- NeXus → Using purl.org for the ontology and term IRIs.
- PaNET → Using purl.org for the ontology and term IRIs.
- PaN-mapping → Using purl.org for the ontology and no new terms have been created.

22

Ten Simple Rules for Making a Vocabulary FAIR - evaluation

Rule 6. Create machine readable representations of the vocabulary terms
NeXus →
PaNET → all: usage of RDF and OWL encoding
PaN-mapping →

Rule 7. Add vocabulary metadata
NeXus → License, Version, Creators/Contributors, Creation date added.
PaNET → License, Version, Creators/Contributors, Creation date added.
PaN-mapping → License, Version, Creators/Contributors, Creation date added.

Rule 8. Register the vocabulary
NeXus → Might be registered in the NCBO BioPortal to make it findable.
PaNET → The vocabulary has been registered in the NCBO BioPortal to make it findable.
PaN-mapping → not registered.

Rule 9. Make the vocabulary accessible for humans and machines
NeXus → Vocabulary is accessible via the GitHub interfaces and PURL redirection.
PaNET → Vocabulary is accessible via the GitHub interfaces and PURL redirection.
PaN-mapping→ Vocabulary is accessible via the GitHub interfaces and PURL redirection.

Rule 10. Implement a process for maintaining the FAIR vocabulary
NeXus → Evolvement of the ontology is linked to the evolvement of the NeXus standard.
PaNET → Issue templates have been created to enable the community to submit.
PaN-mapping → Github issues can be created for maintanace.

HZB Helmholtz Zentrum Berlin
diamond

23