

OSCARS

Open Science Clusters' Action
for Research & Society

Funded Project

Findable Big Data from Various Material Characterisation Techniques

Presenter: Sandor Brockhauser, MPI CBS, <https://orcid.org/0000-0002-9700-4803>

Implemented by



FAIR Data Infrastructure
for Physics, Chemistry,
Materials Science,
and Astronomy e.V.



HZB Helmholtz
Zentrum Berlin



Funded by
the European Union

FAIRness of Experimental Data

- Findability and Interoperability needs commonly accepted ontology
 - PaNET aims to cover experiment techniques
 - NeXus provides details of experiment data
 - ESRF-ET is to connect experiment data to techniques
 - Data repositories should offer searchability according to these ontologies
 - Data driven science and AI tools need such platforms
-

Main goals

- Ontology mapping:
PaNET - NeXus - ESRF-ET
 - Ontology integration:
RDM platforms (E.g. NOMAD, ESRF_DATA, HZB_DATA)
 - Scientific use case:
AI driven spectroscopy analysis
-

First Milestones - ESRF-ET

- more than 30 techniques from ESRF implemented
 - concepts/building blocks for differentiating the techniques via use of object properties
 - Conclusion:
 - ESRF-ET can be split:
 - generic concepts which can be applied at other facilities, too
 - beamline specific concepts
 - Beamline Ontology is a time-ontology
 - DAQ can integrate ESRF-ET annotation:
 - now: manual selection by users/beamline scientists
 - future: automatic selection based on beamline configuration
-

First Milestones - PaNET

- PaNET is a Taxonomy with multiple inheritance
 - difficult to oversee
 - difficult to maintain or extend
 - difficult to avoid contradictions
 - Our proposal: generic concepts from ESRF-ET could be incorporated
 - PaNET would move from Taxonomy to Ontology
 - object properties should be added
 - reasoners could find inconsistencies
-

First Milestones - NeXus Ontology

- NeXus Ontology implements
 - all NeXus concepts (Groups, Fields, Attributes) and their relationships
 - OWL restrictions
 - required vs. optional
 - enumerations
 - data types
 - unit categories
 - Only 3 object properties: has, hasValueContainer, hasUnitContainer
 - Reasoners are slow.
Potential improvement by SVO language modification:
 - enrich object property set, and maybe reduce concepts
-

Taxonomy of NeXus Concepts

Data properties | Annotation properties | Datatypes | Individuals

Classes | Object properties

Class hierarchy: NeXusObject

- owl:Thing
 - NeXus
 - NeXusDataType
 - NeXusObject**
 - NeXusApplicationClass
 - NeXusBaseClass
 - NeXusGroup
 - NeXusQuantity
 - NeXusAttribute
 - NeXusField
 - NeXusUnitCategory

Annotations Usage

Annotations: NeXusObject

Annotations +

[rdfs:comment](#) [type: xsd:string]

NeXus Object (All the concepts defined by the NeXus definitions)

Description: NeXusObject

SubClass Of (Anonymous Ancestor)

Instances +

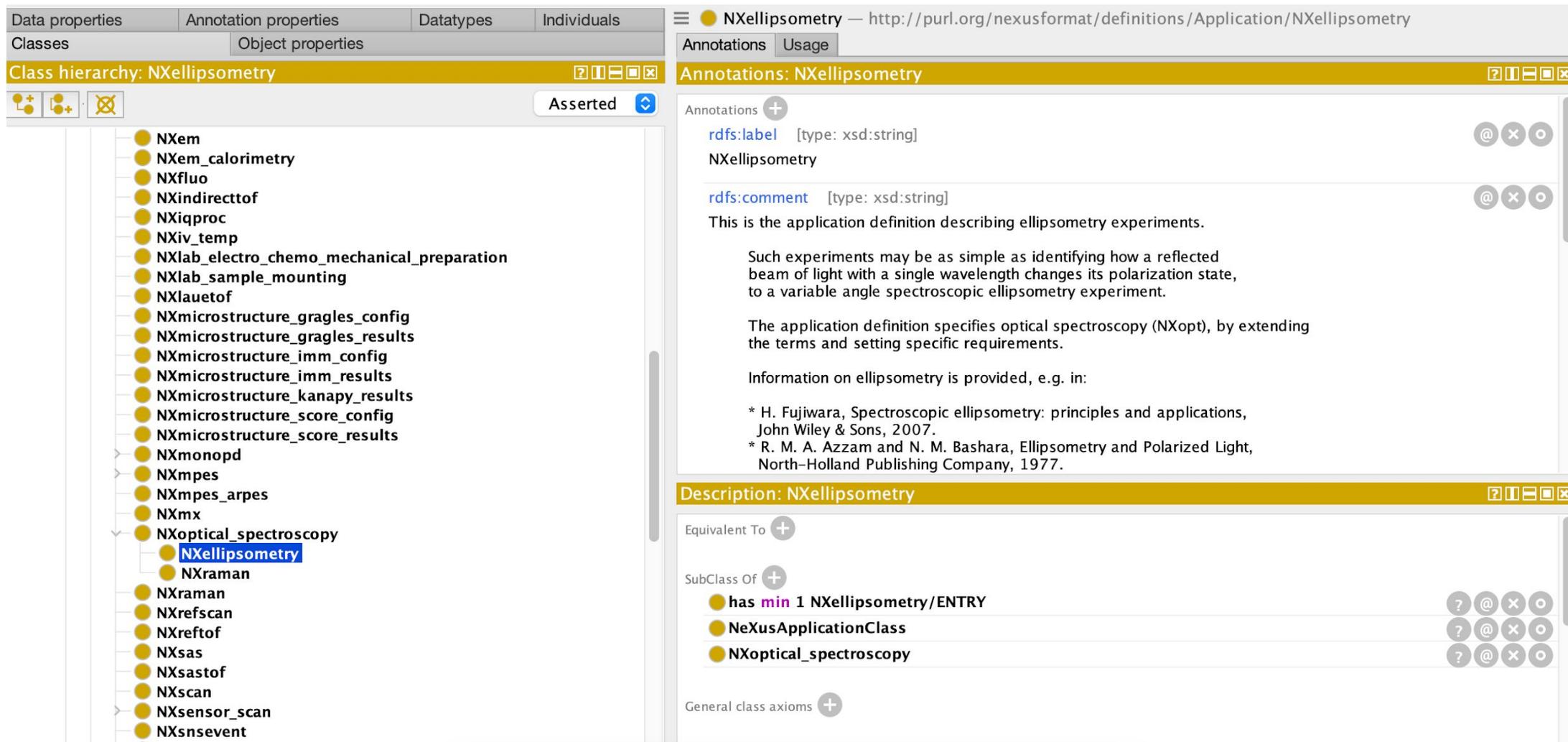
Target for Key +

Disjoint With +

NeXusUnitCategory, NeXusDataType

Disjoint Union Of +

Taxonomy of Application Definitions



The screenshot displays the ontology page for **NXellipsometry** in a web browser. The interface is divided into two main sections: a class hierarchy on the left and a detailed view of the selected class on the right.

Class Hierarchy (Left Pane):

- Classes
- Object properties
- Class hierarchy: NXellipsometry
- Asserted
- Class list:
 - NXem
 - NXem_calorimetry
 - NXfluo
 - NXindirectof
 - NXiiproced
 - NXiv_temp
 - NXlab_electro_chemo_mechanical_preparation
 - NXlab_sample_mounting
 - NXlauetof
 - NXmicrostructure_gragles_config
 - NXmicrostructure_gragles_results
 - NXmicrostructure_imm_config
 - NXmicrostructure_imm_results
 - NXmicrostructure_kanapy_results
 - NXmicrostructure_score_config
 - NXmicrostructure_score_results
 - NXmonopd
 - NXmpes
 - NXmpes_arpes
 - NXmx
 - NXoptical_spectroscopy
 - NXellipsometry** (selected)
 - NXraman
 - NXraman
 - NXrefscan
 - NXreftof
 - NXsas
 - NXsastof
 - NXscan
 - NXsensor_scan
 - NXsnsevent

Class Details (Right Pane):

Annotations: NXellipsometry

- rdfs:label** [type: xsd:string]: NXellipsometry
- rdfs:comment** [type: xsd:string]: This is the application definition describing ellipsometry experiments.
Such experiments may be as simple as identifying how a reflected beam of light with a single wavelength changes its polarization state, to a variable angle spectroscopic ellipsometry experiment.
The application definition specifies optical spectroscopy (NXopt), by extending the terms and setting specific requirements.
Information on ellipsometry is provided, e.g. in:
 - * H. Fujiwara, Spectroscopic ellipsometry: principles and applications, John Wiley & Sons, 2007.
 - * R. M. A. Azzam and N. M. Bashara, Ellipsometry and Polarized Light, North-Holland Publishing Company, 1977.

Description: NXellipsometry

- Equivalent To: +
- SubClass Of: +
 - has **min 1** NXellipsometry/ENTRY
 - NeXusApplicationClass
 - NXoptical_spectroscopy
- General class axioms: +

First Milestones - Use of Ontologies

- Alternatives:
 - (a) external services, but no modification in RDM solutions
 - (b) integrated solutions
 - NOMAD: search for NXoptical_spectroscopy shall also find NXraman, ...
 - Elasticsearch:
 - fast search based on preindexed searchable keywords
 - no meaning and semantics
 - alternative a: search for all sub-classes and merge results
 - alternative b: register all super-classes and have a single ES query
-

Progress Report after 16 months

- ESRF-ET
 - All regular techniques are implemented
 - concepts and object properties for differentiating the techniques
 - PaNET
 - Proposal for improving the ontology and incorporate ESRF-ET elements
 - NeXus Ontology
 - application definitions are connected to PaNET/ESRF-ET
 - Use of Ontologies
 - ESRF solution with DAQ integration
 - NOMAD solution at data injection time
-

Final Steps

- Scientific Use case:
 - Data portals to support PaNET based queries
 - Find and download spectroscopy datasets from different portals
 - Scientific data analysis in NOMAD's Artificial Intelligence Toolkit
-