



# FAIR Ontology Metadata

## Crucial Information for Understanding, Evaluation, and Application

Susanne Arndt\*, Hendrik Borgelt\*\*, Philip Strömert\*\*\*

[2nd Ontologies4Chem Workshop 2023](#) - Day 2. October 12th - 12th 2023

DOI: [10.5281/zenodo.10117899](https://doi.org/10.5281/zenodo.10117899)

\* NFDI4Ing, TIB - Leibniz Information Centre for Science and Technology, Hanover, German  
\*\* NFDI4Cat, Dept. of Biochemical and Chemical Engineering, TU Dortmund University, Germany  
\*\*\* NFDI4Chem, TIB - Leibniz Information Centre for Science and Technology, Hanover, German

# Why We Need an Ontology Metadata Standard



Terminologies used for FAIR RDM need to be FAIR themselves!



- Depends on trusted, reliable hosting & indexing services
  - Usable for humans & machines (API)



- Depends on alignment or mappings between terminologies
  - e.g. shared upper & mid level ontologies, reusing terms



- Depends on open licenses, good documentation & maintenance
  - e.g. CC-BY, active community curation, version controlled

**Ontology metadata play a crucial part in this!**

# Why We Need an Ontology Metadata Standard

---



- **indexing** ontologies in registries and look-up services
  - our primary use case in NFDI is the provision of ontologies in curated sets
- **evaluating** the use case applicability & quality of ontologies
  - applications using ontologies need to display context information
  - deriving information only from the source code can be expensive or even impossible
- **validating** their overall FAIRness and metadata quality
  - although some metadata will always be subjective

We need to have answers to basic questions around

**What? Where? When? Who? Why? How?**

# Which Ontology Metadata are Needed & Where

---



**What** is the unique name and scope of an ontology?

**When** was it created and when last modified?

**Where** does the ontology source code live?

**Who** is responsible for it and who can reuse it?

**Why** is it relevant in certain contexts?

**How** can and should it be reused?



**Problem 1: The answers to these questions must be formalized**

**Problem 2: They should at best be part of the ontology source code**

# Towards an Ontology Metadata Standard

---



Recommendations on what needs to be formalized exist already!

- [LOD](#) & [OBO Foundry](#) best practices
- [Metadata for Ontology Description and Publication Ontology \(MOD\)](#)
  - Reusing standards like Dublin Core and DCAT whenever possible

Use case dependent profiles must be possible

- Cardinality constraints vary by use case, e.g. indexing vs quality analysis
- Equivalent attributes need to be mapped, e.g.:
  - Ontology level: *dct:title* = *dc:title* = *sdo:name*
  - Term level: *rdfs:label* = *skos:prefLabel*

# Towards an Ontology Metadata Standard

## TIB Use Case: Ontology Metadata Needed for Indexing

---



### Recommendations for our TIB & NFDI4Chem Terminology Service

- Still a WIP → ontology level metadata first
  - Automating ontology ingest → parse from source & curate less
- **required:** PURL / file path, title, acronym / prefix, creator, version, creation & modification date, scope description, license, documentation and issue tracker
- **recommended:** contact, contributors, funding, audience, subject, language, syntax, status, source repo, citation & references (incl. derived from) and root terms
- **optional:** abstract, alt. title & prefix, related versions / IDs, homepage, social media, logo, KOS type, example ID, alignments & mappings, competency questions, methodology, dev environment, mailing list, publisher and comments

# Towards an Ontology Metadata Standard

## NFDI4Chem Use Case

---



### Ontology Level Metadata needed in NFDI4Chem

- Scope of an ontology
  - Ability to quickly grasp use case applicability
- Dependencies to other ontologies
- Used in which projects
  - Being reused actively vs being only indexed for research
- Current status & what needs to be done
  - E.g. active, draft, retired/orphaned (see FIX & REX)
  - Issue tracker → new GitHub issue



# Towards an Ontology Metadata Standard

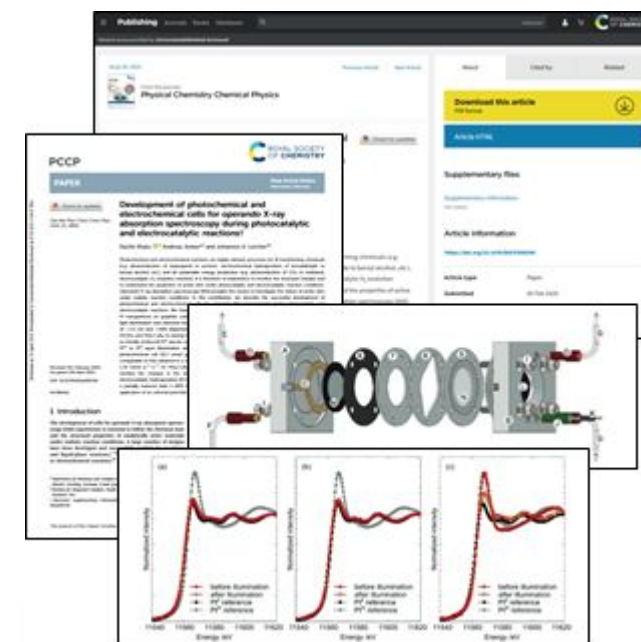
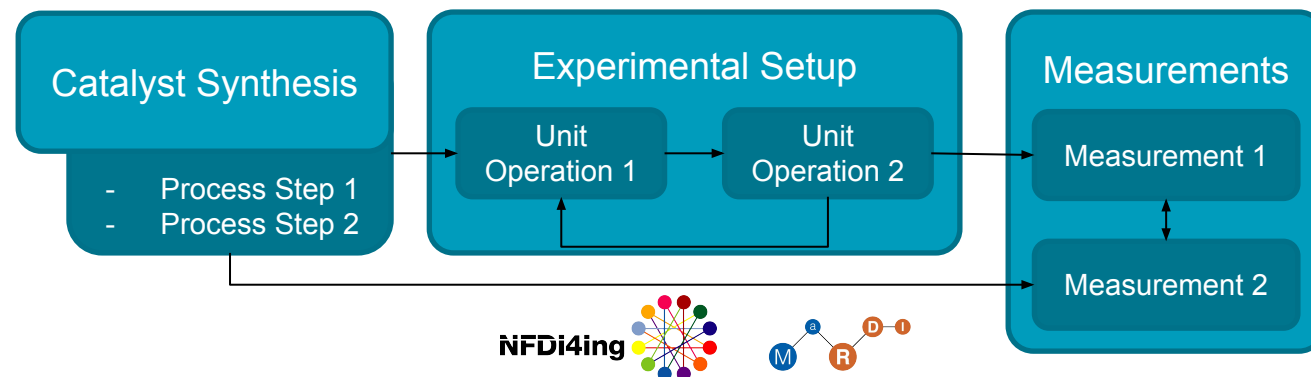
## NFDI4Cat Use Case

### Challenges in Catalysis

- Modelling coupled data
  - Catalyst/chemical provenance
  - Processes & parameters
- Few relevant cross domain ontology level metadata terms
  - Too generic vs too specific

### Bridging to Ontology Level Metadata needed in NFDI4Cat

- Metadata framework required
  - Catalysis researcher must be able to understand this easily
  - Many domains to be mapped
- Current status & what needs to be done
  - Further term & ontology search



[1] <https://pubs.rsc.org/>

[2] Khare, Rachit; Jentys, Andreas; Lercher, Johannes A. (2020): Development of photochemical and electrochemical cells for operando X-ray absorption spectroscopy during photocatalytic and electrocatalytic reactions. DOI: 10.1039/d0cp00654h.

[1,2]



# How to Validate Ontology Metadata?

---



## SPARQL approach

- [ROBOT report](#)
  - Used within the Ontology Development Kit
  - Default is mostly [term level](#), only *title*, *description* & *licence* on the ontology level

## SHACL based approach

- [Ontology Access Kit](#) - SHACL derived from a (quite complex) [LinkML schema](#)
  - Planned to be customizable with own schema → idea of different schema profiles
- DCAT-AP based [Metadata Quality Assessment \(MQA\)](#) tool → not ontology specific enough

## Application logic based approach

- e.g. [OLS](#), [OBO Foundry registry](#), [Bioregistry](#), OntoPortal's [O'FAIRe](#), [TIB & NFDI4Chem TS](#)

# How to Validate Ontology Metadata?

## TIB/NFDI4Chem Use Case: Our Validation Approaches



### SHACL shapes approach

- Manually written for now
- Testing with a variety of already indexed ontologies
- Intended as part of our automated ingest pipeline
- Logging levels: error, warning, info
  - Helping the user when indexing request feature is implemented

```
# Start of the Ontology Metadata Shape
ontom:OntologyMetadataShape a sh:NodeShape ;
  sh:targetClass owl:Ontology ;
# MANDATORY METADATA
## G-Doc https://docs.google.com/document/d/1FRmCQ7eD6PcqSQgPbqCGjki3t39R0oYozsehCPX9E0Q/
edit#bookmark=id.w5pmqfs8xiuv

## Ontology title
### G-Doc https://docs.google.com/document/d/1FRmCQ7eD6PcqSQgPbqCGjki3t39R0oYozsehCPX9E0Q/
edit#bookmark=id.1exx8hu4osgx
  sh:property [
    sh:path [sh:alternativePath (rdf:type skos:prefLabel dc:title dcterms:title schema:name) ] ;
    sh:severity sh:Violation ;
    sh:minCount 1 ;
    sh:message "The ontology does not specify its label or title."@en ;
    sh:description "The ontology must specify its label or title."@en ;
    sh:name "Title"@en ; # this is for forms with DASH
    sh:order 1 ; # also for forms, order of fields
  ] ;

  sh:property [
    sh:path [sh:alternativePath (rdf:type skos:prefLabel dc:title dcterms:title schema:name) ] ;
    sh:severity sh:Info ;
    sh:datatype rdf:langString ;
    sh:message "The ontology title is not an rdf:langString, i.e. it does not carry a language tag.
      recommend to add one."@en ;
    sh:description "The ontology title must be an rdf:langString, i.e. it does not carry a language
      tag."@en ;
    sh:name "Title"@en ; # this is for forms with DASH
    sh:order 1 ; # also for forms, order of fields
  ] ;
```

### LinkML approach (testing)

- Based on exchanges with Chris Mungall and Nico Matentzoglou
  - Vision: TIB profile as an extension of a more common base profile
- Could produce needed shapes automatically → TODO equivalent slot URIs mappings
- Benefit of multiple representations → other validation frameworks might be better than SHACL



# How to Validate Ontology Metadata?

## NFDI4Cat Use Case: Validation and Implementation Approaches

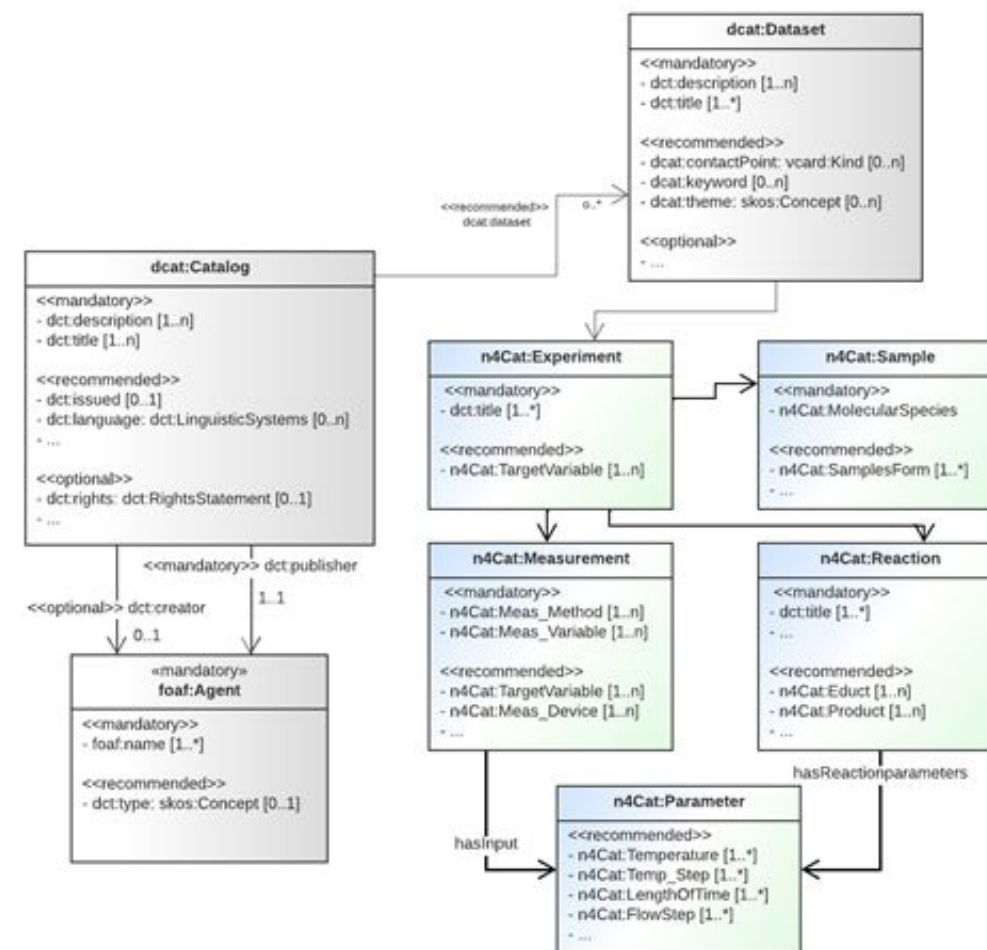


### DCAT-AP SHACL shapes with LinkML

- LinkML as simplified language
- Term reuse from schema, bioschemas, Voc4cat

### Current status & what needs to be done

- Metadata framework development
  - LinkML integration (LinkML to SHACL)
- Terms search and domain differentiation
  - Use case based approach
  - SHACL based mapping tool development



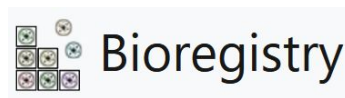
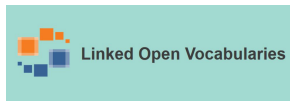
# Sustainability & Community Curation of Ontology Metadata



What do we do when needed metadata is not part of the ontology source code?

→ **It needs to be manually curated somewhere!**

- Already happens in multiple places, e.g.:



- Plurality is a good thing → usability → redundancy / resilience → sustainability
- Curation efforts just need to be bundled, shared and done collaboratively
  - Saves resources and fosters a healthy open science community rdfs:seeAlso [great\\_choyt\\_talk](#)
  - **Based on normalized metadata slots + different cardinality profiles**
    - LinkML ? → multiple representations (SPARQL, SHACL, JSON-LD, Pydantic, Docs ...)
- Enables easier upstream pull requests → better ontologies

