Common Minimum Metadata for FAIR Semantic Artefacts

Clement Jonquet, Biswanath Dutta, Luiz O. Bonino da Silva Santos, Robert Pergl, Yann Le Franc

Common Minimum Metadata for FAIR Semantic Artefacts

Clement JONQUET, [a,b,1] https://orcid.org/0000-0002-2404-1582
Biswanath DUTTA, [c] https://orcid.org/0000-0003-3059-8202
Luiz O. BONINO DA SILVA SANTOS, [d,e] https://orcid.org/0000-0002-1164-1351
Robert PERGL, [f,g] https://orcid.org/0000-0003-2980-4400
Yann LE FRANC [f,1] https://orcid.org/0000-0003-4631-418X
[a] *LIRMM, University of Montpellier & CNRS, France*
[b] *MISTEA, University of Montpellier, INRAE & Institut Agro, France*
[c] *DRTC, Indian Statistical Institute, Bangalore, India*
[d] *Semantics, Cybersecurity & Services, University of Twente, The Netherlands*
[e] *Biosemantics group, Leiden University Medical Center, The Netherlands*
[f] *e-Science Data Factory, France*
[g] *Faculty of Information Technology, CTU in Prague, Czech Republic*

Abstract.
Semantic interoperability is crucial for the FAIR Principles and strongly relies on Semantic Artefacts that also need to be FAIR. To achieve this, semantic artefacts require rich, structured, and interoperable metadata. The challenge lies in determining the threshold for "rich metadata" and agreeing on a common minimum set. The H2020 FAIRsFAIR project and the RDA *Vocabulary Semantic Services Interest Group* addressed this question by developing a "minimal metadata model" for semantic artefacts. In this paper, we present background information, methodology, discussions and workshops which contribute to the establishment of the *FAIRsFAIR minimum metadata profile for semantic artefacts.* We present an extension of the Metadata for Ontology Description and Publication Ontology (MOD2.0) incorporating this profile as well as its implementation (SemanticDCAT-AP) and its use to build FAIRcat, a prototype of a FAIR Data Point harvesting the content of multiple semantic artefact catalogues.

Keywords. Minimal metadata model, Semantic artefacts, Ontologies, Vocabularies, Metadata, FAIR, FAIRness assessment.

1. Introduction

Semantic interoperability is at the very core of the FAIR Data Principles [1] and as in any interoperability effort, it requires agreement on how the resources or artefacts supporting "semantics" are described. In all domains, many vocabularies, terminologies, ontologies or more largely semantic artefacts[2] are produced to represent and annotate data to make them more interoperable. Semantic artefacts have even become a master element to achieve FAIRness and have been discussed as digital objects that themselves need to be FAIR [2–4].

[1] Corresponding authors: C. Jonquet (jonquet@lirmm.fr), Y. Le Franc (ylefranc@esciencefactory.com).

[2] Semantic artefact is a broader term, originally proposed in [2], and more and more used to include ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas and standards.

However, in order to properly follow the FAIR principles, semantic artefacts need rich, structured and interoperable metadata, which is also a necessary condition for machine-actionability [1]. One of the main challenges for implementing the FAIR principles, whether for semantic artefacts or for any kind of data, is to determine the threshold for "rich metadata" mentioned in principle F2. It requires communities to agree on a common and minimal metadata schema that could be used as a threshold for FAIR. Reaching an agreement on such a common metadata schema and representation improve systems' interoperability by allowing the development of client applications that would need to read/parse only one representation format. On top of this, by agreeing on a minimum set of metadata for semantic artefacts, the same systems would guarantee every time an application encounters the metadata of a semantic artefact, it would know that minimally certain information would be available.

The task 2.2 of the H2020 FAIRsFAIR project was dedicated to establish prerequisites for better semantic interoperability by developing recommendations and facilitating uptake of good practices to make semantic artefacts compliant with the FAIR principles [2–4]. Among the recommendations produced by the project, *P-Rec 3* required the creation of "*A common minimum metadata schema* to *be used to describe semantic artefacts and their content."* Such a minimum set of metadata (also called a "minimal metadata model") for semantic artefact was missing. To fill in this gap, the task established, in collaboration with a wide range of communities, a first version of this model presented here. This collaboration was supported by the RDA Vocabulary Services Interest Group's (VSSIG) Ontology Metadata task group which was discussing an extension of the pre-proposed *Metadata for Ontology Description and Publication Ontology* (MOD) model [5]. This extension would both: (i) offer a review of all metadata properties available for semantic artefacts (similar to a "maximal metadata model") and (ii) revise the MOD model as an extension of DCAT2.

This joint effort had the overall goal to enable the implementation of FAIRness assessment methods that would be capable of establishing some kind of base threshold for a semantic artefact to be FAIR, as for example the grid proposed in [6]. In doing this, one challenge was to identify both: (i) generic metadata properties for digital objects that would apply to semantic artefacts (e.g., creator, identifier, or license); and (ii) metadata properties that would be specific to semantic artefacts (e.g., representation language). Furthermore, a shared understanding has to be reached on the value, necessity and feasibility of the key metadata properties. Indeed, MOD v1.4, released in 2018, contained 128 properties to describe semantic artefacts. Those were taken from 15 "crosswalked" metadata vocabularies[3] and would come as this, without prioritization, so that a developer would face a concrete problem of identifying which are the key/most important of these properties. Of course, nothing would restrict the usage of more metadata properties for richer description.

Therefore, the subject of minimal metadata model for semantic artefact was discussed openly and publicly in multiple workshops and meetings organized by the FAIRsFAIR task 2.2 and RDA VSSIG task groups. Finally, a workshop organized June 4th 2021 –in which more than 30 participants from around 20 different communities contributed (out of 76 attendees)– helped us to come to the *FAIRsFAIR minimum metadata profile for semantic artefact* presented in this article.

[3] By "crosswalked", we mean that we have identified 346 metadata properties in those metadata vocabularies that, once mapped (crosswalks identified), bring us to 128.

In the following, we present background information on the subject of FAIR semantic artefacts (Section 2), then we introduce our working methodology (Section 3). We briefly present the MOD2.0 proposed model for semantic artefact and their catalogues done by extending DCAT2 (Section 4).[4] Then, based on the two new main objects introduced by this model, `mod:SemanticArtefact` and `mod:SemanticArtefactDistribution`, we present the metadata properties that were gathered to describe them and eventually voted to decide the level of requirement: Mandatory or Recommended or Optional (Section 5). Then, we explain how we have integrated the requirements in MOD2 and also developed two machine-actionable representations of SemanticDCAT-AP, an experimental application profile used by FAIRcat, a tool to aggregate and align repository metadata content to DCAT for publication in a FAIR Data Point (Section 6). Finally, we conclude and present some perspectives (Section 7).

2. Background

Before the FAIR Principles, a recommendation for publishing RDF vocabularies was produced in 2008 by the W3C Semantic Web Deployment Working Group.[5] Then in 2014, the 5-stars LOD principles of Berners-Lee [7] were specialized for linked data vocabularies [8] as five rules to follow for creating and publishing "good" vocabularies. The degree to which the FAIR principles align and extend the 5-star open data principles was also later in studied [9, 10] and [6] presented after. In 2017, the *Minimum Information for Reporting an Ontology* initiative published the MIRO guidelines for ontology developers when reporting an ontology in scientific reports [11]. These guidelines refer to 34 information items (such as "ontology name," "ontology license," "ontology URL") and specify the level of importance (must, should, optional) for each individual information item. This work was significant but was never put in perspective with the FAIR principles. In MOD, where the authors reviewed which properties of MOD v1.4 could "help" addressing which MIRO guidelines (cf. example in section 6.1).

In 2017, Dutta et al. [5] reviewed and harmonized existing metadata vocabularies and proposed a unified *Metadata for Ontology Description and Publication Ontology* (MOD) model to facilitate manual and automatic ontology descriptions, identification, and selection. MOD is not another standard nor another metadata vocabulary, but more an aggregated set of identified properties one can use to describe a semantic resource.[6] MOD 1.4 was used in AgroPortal to implement a richer, unified metadata model [12].

Then, since 2020, we have seen four parallel initiatives that investigated the question of FAIR semantic artefacts:

- In March 2020, the FAIRsFAIR H2020 project delivered the first version of a list of 17 recommendations and 10 best practices recommendations for making semantic artefacts FAIR [2]. For each recommendation, the authors provided a detailed description associated with a list of related supporting technologies or technical solutions proposed by different communities.

[4] This model is still being consolidated (now in the context of the Horizon Europe FAIR-IMPACT project: https://github.com/FAIR-IMPACT/MOD), but the minimal model can be presented here independently.

[5] https://www.w3.org/TR/swbp-vocab-pub

[6] For instance, MOD does not require the use of a specific authorship property but rather encodes that `dc:creator`; `schema:author`, `foaf:maker`, or `pav:createdBy` can be used to say so.

- Later in 2020, Garijo et al. [13] produced "guidelines and best practices for creating accessible, understandable and reusable ontologies on the Web." In another position paper, Poveda et al. [14] completed their work with a qualitative analysis of how four ontology publication initiatives cover the foundational FAIR principles. They proposed some recommendations on making ontologies FAIR and listed some open issues that might be addressed by the semantic Web community in the future. In October 2021, Garijo et al. proposed FOOPS! a Web service for assessing an ontology regarding the FAIR principles [15].
- Late 2020, Cox et al. proposed guidelines ("10 simple rules") for making a vocabulary FAIR (https://fairvocabularies.github.io/makeVocabularyFAIR) and transform vocabularies that are not available following Web standards [16]. However, the authors do not explain how the proposed rules are aligned to each individual FAIR principle.
- A list of functional metrics and recommendations for *Linked Open Data Knowledge Organization Systems* (LOD KOS) was proposed in 2020 [17].
- In the end of 2020, DBPedia Archivo [18], an ontology archive, was released to help developers and consumers to implement FAIR ontologies. The prototype automatically discovers, downloads, archives, and rates new ontologies (https://archivo.dbpedia.org). Unfortunately, this work had not been inspired by existing research methodologies/tools.
- In 2021, Amdouni et al., introduced an "integrated quantitative FAIRness assessment grid for semantic resources [6]. This work was nourished and aligned with relevant state-of-the-art initiatives for FAIRness assessment: the RDA FAIR Data Maturity Model, the RDA Sharing Rewards and Credit evaluation table, the 5-stars for vocabulary as well as FAIRsFAIR and Poveda et al. recommendations cited above. The grid dispatches different credits to each FAIR principle, depending on its importance –according to pre-existing initiatives– when assessing FAIRness.
- Early 2022, the same authors proposed a metadata-based automatic FAIRness assessment methodology for ontologies and semantic resources called *Ontology FAIRness Evaluator* (O'FAIRe), based on the grid described previously [19]. The methodology projects the 15 foundational FAIR principles for ontologies, and proposes 61 questions, among which 80% are based on the resource metadata descriptions. The methodology has been (partially) implemented in AgroPortal [20] and is currently being transferred to other OntoPortal-based ontology repositories.

In conclusion, each of these initiatives reviewed somehow –more or less directly– some metadata properties associated to multiple criteria required to produce a so-called "FAIR semantic artefact"; however, none of these approaches explicitly list a minimal set of metadata properties to be considered FAIR and took the responsibility to qualify these properties as mandatory-recommended-optional. We believed: (i) a consensual approach, based on informal feedback and voting was actually a good way to converge and (ii) our current research projects and working groups were offering a relevant context for discussion such a consensus.

3. Methodology

In 2020-2021, in parallel with the RDA VSSIG Ontology Metadata Task Group which was working on defining MOD2 presented in Section 4, the H2020 FAIRsFAIR project organized three public workshops to eventually produce the minimal model presented in

Section 5. These workshops involved ontologists, knowledge engineers and semantic artefact catalogue providers:

- On April 29th, 2020 (~30 participants): the objective was to present and discuss the first set of 17 "general recommendations" and 10 "best practices recommendations" for FAIR semantic artefacts [2] and in particular about P-Rec3 on metadata for FAIR Semantic artefacts. The recommendations were also made publicly available for comments on GitHub.[7] The outcomes of this workshop as well as the discussion on GitHub and in subsequent RDA task groups meetings contributed to revise the recommendations and produce a second version [3]
- On October 15th, 2020 (~30 participants): the objective of this second workshop was to collect feedback on the first version of the recommendations and to establish the alignment of the recommendations with the RFC 2119 (MUST, SHOULD, SHALL). The outcomes of this workshop contributed to the second version of the recommendations [3].
- On June 4th, 2021 (~80 participants): the objective was then to determine a set of key metadata properties to build a minimum metadata profile for semantic artefacts, setting up a threshold on FAIRness. In this workshop, the participants voted to decide if each property should be optional, recommended or mandatory. During the votes the participants focused on the meaning of the properties i.e., the information they encode, but not necessarily on the metadata vocabularies providing a formal property to encode this information. At the beginning of the workshop, all attendees were made aware of the idea of data modeling and a good familiarity with DCAT was suggested as these were needed to actually contribute during the voting session. They were then presented with a simple use-case to support the voting: *what would be the necessary fields for retrieving semantic artefacts?* Both DCAT and the MOD2 proposition were thoroughly presented, then workshop participants were asked only to share responses for which they considered themselves to have sufficient expertise and awareness to make an informed contribution. They were guided through this process by the organizers (authors).

4. A proposed model for semantic artefact and their catalogues (MOD2)

MOD2.0 was proposed in 2020 as a new version of the *Metadata for Ontology Description and Publication Ontology,* structured as an extension of DCAT.[8] The Data Catalog Vocabulary (DCAT) is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. By using DCAT to describe datasets in catalogues, publishers increase discoverability and enable applications to consume metadata from multiple catalogues. The key idea in extending DCAT was to view semantic artefacts as "datasets of knowledge entities" that can be available in multiple "distributions" and can be "cataloged" in repositories such as BioPortal [21] or the Ontology Lookup Service [22].[9]

In designing MOD2, several design issues –not necessarily discussed in this paper– were raised including: (i) how (and is it necessary) to "extend" the notion of distribution?;

[7] https://github.com/FAIRsFAIR-Project/FAIRSemantics/issues

[8] MOD2 proposition was released in 2020 on GitHub and presented in multiple workshops and talks. However, the new model made of the 5 new classes presented here has never been published in a scientific communication yet.

[9] We now use the expression "Semantic Artefact Catalogue".

(ii) which classes inside DCAT and outside, the `mod:SemanticArtefact` class should explicitly extend or supersede?; (iii) which metadata properties from other vocabularies are available to describe semantic artefacts?; (iv) are there properties from outdated and not maintained metadata vocabulary that MOD could adopt?. The five key-classes from DCAT were finally specialized by creating new classes in the MOD namespace, as illustrated in Figure 2:

- `mod:SemanticArtefact`: A collection of knowledge entities (classes, properties, concepts, terms, mappings), produced and curated by a single or multiple agents, and available for access or download in one or more representations. This is typically the class of any knowledge organization systems or resources such as ontologies, vocabularies, concepts schemes, thesauri, terminologies, etc. For example, the AGROVOC thesaurus (http://aims.fao.org/aos/agrovoc) or the CODO ontology (https://w3id.org/codo).
- `mod:SemanticArtefactDistribution`: A specific representation of a semantic artefact. Typically, the class of any possible distributions or issuances of the semantic artefacts. It could be used to distinguish either multiple versions of a semantic artefact or different format/representation available. For example, "the version 1.3 of CODO in OWL"; or the "AGROVOC Core distribution in SKOS with TTL syntax".
- `mod:SemanticArtefactCatalog`: A curated collection of metadata about semantic artefacts. Typically, the class of repositories, libraries or services hosting and maybe also serving various semantic artefacts. For example, the NCBO BioPortal repository or the AgroPortal vocabulary and ontology repository or the NERC Vocabulary Server.
- `mod:SemanticArtefactCatalogRecord`: A record in a catalog, describing the registration of a single semantic artefact. Typically, the class of the entries for semantic artefacts inside catalogues i.e., when a catalogue hosts a semantic artefact, it is often concretely materialized by a record describing the artefact following the catalogue metadata model. For example, the record for CODO in BioPortal (https://bioportal.bioontology.org/ontologies/CODO) or the record for AGROVOC in AgroPortal (http://agroportal.lirmm.fr/ontologies/AGROVOC).
- `mod:SemanticArtefactService`: A collection of operations providing access to one or more semantic artefacts or SemanticArtefact-based processing functions/services. Typically, the class of the services offered for semantic artefacts. For example, the REST API of BioPortal, the SPARQL endpoint of AgroPortal, or the browsing user interface of a SKOSMOS based service, a FAIR data point.

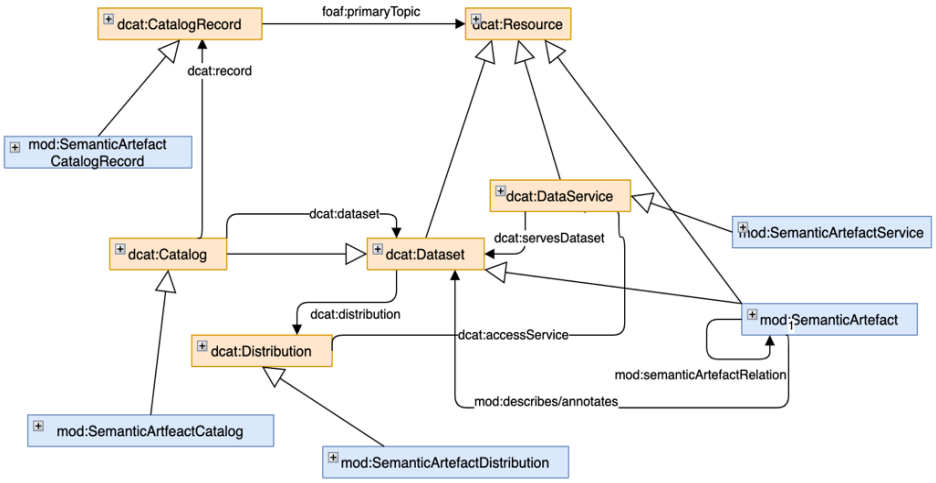

Figure 1. MOD2 proposed model for semantic artefact and their catalogues.

By inheriting from `dcat:Resource` (itself an `rdfs:Resource`) and `dcat:Dataset`, the `mod:SemanticArtefact` and the mod:`SemanticArtefactDistribution` classes could be described by the properties compatible with and suggested by DCAT. Plus, we decided to adopt and generalize in the MOD namespace the properties from OMV [23], DOOR [24] and VOAF (http://purl.org/vocommons/voaf) as those vocabularies were very specific to ontologies/vocabularies but are not maintained anymore.[10] Finally, we also had to re-incorporate the metadata properties from the previous versions of MOD (v1.4) to the relevant class (SemanticArtefact or Distribution), as MOD v1.4 did not distinguish the two. These three steps gave us a set of 92 properties (for `mod:SemanticArtefact`) and 46 properties (for `mod:SemanticArtefactDistribution`) from which we selected respectively a subset of 41 (Table 1) and 24 (Table 2) properties for the voting workshop. We have not yet status on the properties for the other new classes created (Service, Catalog, CatalogRecord) but we anticipate specialized properties justifying the creation of a new subclass in MOD2.

5. FAIRsFAIR Minimum metadata recommendations for semantic artefact

After the workshop, the results were collected into a spreadsheet for evaluation. The inputs were the number of votes for each endorsement level mandatory-recommended-optional. From them, we calculated the percentage of votes for each option. The option with the highest percentage was then selected. In cases where two options voted the same, the third was taken into account, e.g., 46.43% mandatory, 46.43% recommended, 7.14% optional would imply recommended, as 7.14% also voted for optional. As an auxiliary informative metric, we computed the consensus of the voting using the following formula:

consensus = (0.333 + *percentage of votes for the winning option* − *sum of percentages of the non-winning options*) / 1.333.

[10] Furthermore, we were motivated by unifying all the metadata properties *specific* to semantic artefacts.

Table 1 and Table 2 shows the result for all the properties voted.

Table 1. List of properties, and decision for the `mod:SemanticArtefact` class.

Property	# of votes	Consensus	Decision
dct:title	**23**	86.95%	mandatory
dct:license	**28**	78.57%	mandatory
dct:identifier	**28**	67.85%	mandatory
dct:accessRights	**31**	66.12%	mandatory
dct:creator	**31**	66.12%	mandatory
dct:created	**13**	53.83%	mandatory
dct:description	**28**	51.77%	mandatory
dcat:contactPoint	**31**	46.76%	mandatory
owl:versionIRI	**13**	42.29%	mandatory
dct:modified	**28**	35.70%	mandatory
dcat:keyword	**31**	32.24%	mandatory
mod:acronym	**13**	30.75%	mandatory
dcat:landingPage	**31**	22.56%	mandatory
dct:publisher	**28**	19.62%	recommended
dct:subject	**13**	19.21%	mandatory
dct:type	**22**	18.16%	mandatory
dct:issued	**27**	16.65%	mandatory
dcat:theme	**30**	9.98%	mandatory
dct:conformsTo	**30**	14.98%	recommended
dct:language	**28**	19.62%	recommended
mod:URI	**13**	19.21%	optional
dcat:distribution	**25**	15.98%	recommended
dct:contributor	**13**	30.75%	recommended
dct:rights	**23**	34.77%	recommended
dct:temporal	**24**	6.23%	recommended
dcat:qualifiedRelation	**29**	27.57%	optional
mod:status	**13**	42.29%	recommended
odrl:hasPolicy	**23**	21.72%	optional
prov:qualifiedAttribution	**23**	41.29%	optional
prov:wasGeneratedBy	**26**	13.44%	optional
dct:relation	**22**	18.16%	optional
dct:isReferencedBy	**28**	14.26%	optional
schema:includedInDataCatalog	**12**	37.48%	optional
mod:competencyQuestion	**13**	65.38%	optional
dct:accrualPeriodicity	**25**	21.98%	optional
dct:spatial	**24**	18.73%	optional
mod:usedEngineeringMethodology	**12**	24.98%	recommended
dcat:temporalResolution	**25**	39.98%	optional
mod:hasFormalityLevel	**13**	19.21%	recommended
dcat:spatialResolutionInMeters	**25**	57.99%	optional
dct:accrualMethod	**13**	42.29%	recommended

Table 2. List of properties, and decision for the `mod:SemanticArtefactDistribution` class.

Property	# of votes	Consensus	Decision
dcat:mediaType	**17**	47.05%	mandatory
dct:format	**17**	47.05%	mandatory
dct:title	**15**	39.98%	mandatory
dcat:accessURL	**17**	38.22%	mandatory
mod:hasRepresentationLanguage	**10**	24.98%	mandatory
mod:hasSyntax	**10**	24.98%	mandatory
dct:accessRights	**17**	20.57%	mandatory
dcat:downloadURL	**17**	20.57%	recommended

dct:rights	**17**	20.57%	recommended
dct:description	**17**	20.57%	recommended
dct:issued	**17**	2.92%	recommended
dct:modified	**16**	34.36%	recommended
mod:definitionProperty	**10**	39.98%	recommended
dcat:accessService	**16**	24.98%	recommended
dcat:packageFormat	**17**	20.57%	optional
dct:conformsTo	**17**	29.39%	recommended
mod:usedEngineeringTool	**10**	24.98%	optional
mod:prefLabelProperty	**10**	54.99%	recommended
mod:synonymProperty	**10**	24.98%	recommended
odrl:hasPolicy	**15**	39.98%	recommended
dcat:compressFormat	**17**	55.87%	optional
dcat:temporalResolution	**17**	64.70%	optional
dcat:byteSize	**17**	55.87%	optional
dcat:spatialResolutionInMeters	**17**	82.35%	optional

6. Results and applications

6.1. Inclusion of the metadata property requirements in MOD2

We included the requirements within MOD2 as additional information about a metadata property. With this, MOD encodes now three influential works motivating the presence of a property within the vocabulary: (i) the MIRO guidelines followed with using the property; (ii) the FAIR Principle addressed with using the property; and now (iii) the requirement in the FAIRsFAIR profile. For instance, in MOD2, the property `mod:acronym` is encoded as follow:[11]

```
### https://w3id.org/mod#acronym
mod:acronym
   rdf:type                owl:DatatypeProperty ;
   rdfs:subPropertyOf      rdfs:label ;
   rdfs:label              "acronym"@en ,
                           "acronyme"@fr ;
   rdfs:domain             mod:SemanticArtefact ;
   rdfs:range              xsd:string ;
   dcterms:description     "MOD: Short acronym label, often used as an
                           identifier within some ontology platforms such
                           as BioPortal or OBO Foundry. OMV: A short name
                           by which an ontology is formally known."@en ;
   rdfs:isDefinedBy        <http://omv.ontoware.org/2005/05/ontology> ;
   dcterms:issued          "2009-12-24"^^xsd:date ;
   dcterms:relation        <http://www.isibang.ac.in/ns/mod/1.0/acronym> ;
   pav:derivedFrom         <http://www.isibang.ac.in/ns/mod/1.0> ;
   pav:importedOn          "2015-08-05"^^xsd:dateTime ;
   skos:historyNote        "This property has been adopted from OMV
                           Ontology Metadata Vocabulary and redefined in
                           the MOD namespace."@en ;
   prov:wasInfluencedBy    "MIRO guidelines: A.1" ,
                           "FAIR principle: F2" ,
                           "FAIRsFAIR profile: MANDATORY" .
```

[11] https://github.com/FAIR-IMPACT/MOD/blob/master/mod-v2.0_profile.ttl

6.2. *Example of an ontology described with some mandatory metadata*

```
http://myontologyIRI.org
   rdf:type                owl:Ontology; mod:SemanticArtefact ;
   dcat:distribution       http://myontologyIRI.org/distribOWL ,
                           http://myontologyIRI.org/distribPDF ;
   mod:acronym             "MYON" ;
   dcterms:title           "My ontology" ;
   owl:versionIRI          <http://myontologyIRI.org/v1.0> ;
   dcterms:identifier      "myontologyDOI" ;
   dcterms:license         <https://creativecommons.org/licenses/by/3.0> ;
   dcat:landingPage        "myontologyWebPageURL" ;
   dcterms:creator         "http://orcid.org/0000-0002-2404-1582" ;
   dcterms:created         "2023-05-01"^^xsd:dateTime ;
   dcterms:modified        "2023-07-15"^^xsd:dateTime .


http://myontologyIRI.org/distribOWL
   mod:hasRepresentationLanguage  <http://www.w3.org/2002/07/owl>;
   mod:hasSyntax                  <http://www.w3.org/ns/formats/RDF_XML>;
   dcterms:description            "Distribution of My Ontology in OWL";
   dcat:accessURL                 "myontologycataloguerecordURL" .
```

6.3. *SemanticDCAT-AP and FAIRcat*

The minimum metadata profile described in Section 5 have been encoded into RDF/OWL (prefix `semdcat`) [4] (Figure 2). This enables retrieval via simple SPARQL queries and also adding meta-properties (annotations). These meta-properties are:

- `rdfs:definedBy` – the object is a predicate that can be used to retrieve definition of the property. This is required because individual vocabularies employ different approaches e.g., `skos:definition` or `rdfs:comment`.
- `semdcat:endorsement` – the endorsement level being `Mandatory`, `Recommended`, `Optional`. While it can be argued that "mandatory" can be alternatively expressed by OWL axioms, there is no way to express "recommended" without this extension.

At the same time, this approach leads to a more complicated and non-standard RDF/OWL representation, as Object Properties can map only OWL Classes. As such, every property needs an extra class wrapper to be defined.

We also explored an alternative way as SHACL[12] representation of the minimum metadata profile. SHACL shapes are an established way of specifying RDF graph constraints and are extensively used such as in the FAIR Data Point specification [23].[13] This approach allows elegant and straightforward specification of the set of properties, however there are two limitations:

1. Definitions cannot be linked to properties, they must be copied into `sh:description`.
2. Again, there is no way to express "mandatory". The approach taken was to extend the `sh:PropertyShape` with possibility to include `sh-e:endorsement` predicate. As SHACL is RDF, it is formally possible, however such a SHACL file will not pass the standard "SHACL of SHACL" validity, which may be a problem for some checking tools.

[12] https://www.w3.org/TR/shacl
[13] https://specs.fairdatapoint.org

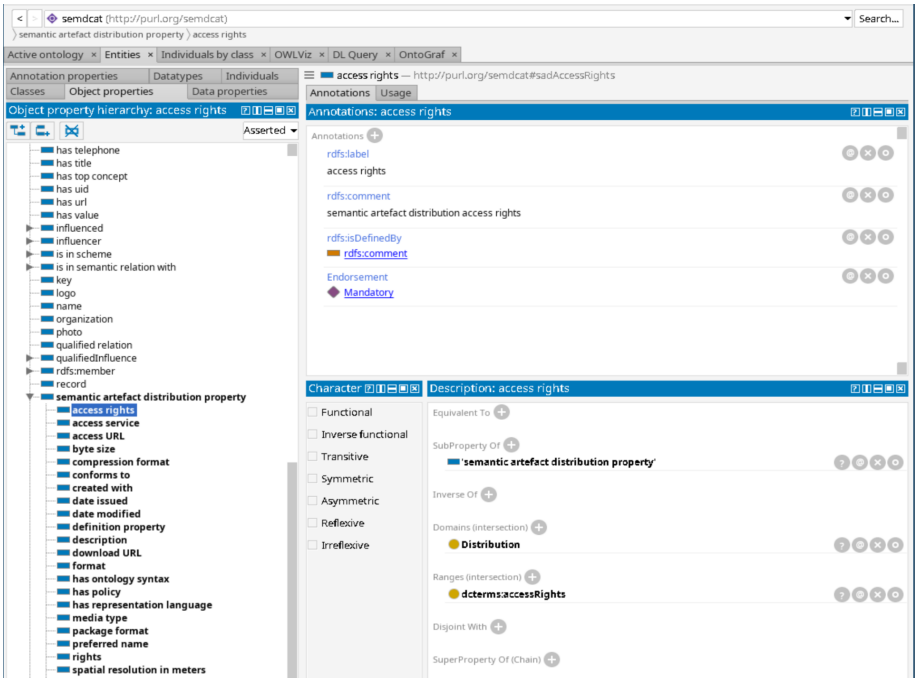

Figure 2. SemanticDCAT-AP's OWL representation in Protégé illustrating the annotations of properties.

An example of the SHACL specification is shown in Figure 3.

```
:SemanticArtefact a sh:NodeShape ;
  sh:name "Semantic Artefact" ;
  sh:description "Semantic Artefact minimum metadata" ;
  sh:targetClass mod:SemanticArtefact ;
  sh:property [
    sh:path dct:accessRights ;
    sh:class dct:RightsStatement ;
    sh:name "access rights" ;
    sh:description "Information about who access the semantic artefact or an indication of its security status." ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
  ], [
    sh:path dct:accrualMethod ;
    sh:name "accrual method" ;
    sh:description "The method by which items are added to a collection." ;
  ], [
    sh:path dct:conformsTo ;
    sh:name "conforms to" ;
    sh:description "An established standard to which the semantic artefact conforms." ;
    sh:nodeKind sh:IRI ;
    sh:maxCount 1 ;
    sh-e:endorsement sh-e:Recommended ;
  ], [
```

Figure 3. SemanticDCAT-AP SHACL definition example. 3rd property illustrates the endorsement extension.

FAIRcat [4] is a proof-of-concept application, based on the federated FAIR Data Space,[14] that utilizes the described OWL-based machine-actionable SemanticDCAT-AP representation and demonstrates the potential of the common minimum metadata for FAIR semantic artefacts. At the same time, it represents a solution that can be used to increase FAIRness of semantic artefacts without any time and resource investments at

[14] https://www.eosc-pillar.eu/federated-fair-data-space-f2ds

the side of repository providers. The idea of FAIRcat is depicted in Figure 4. Semantic artefact catalogues were harvested for their items metadata. Using mappings, they are converted into the SemanticDCAT-AP and stored into a FAIR Data Point, the FAIRsFAIR Reference FAIR Data Point in our case.[15]

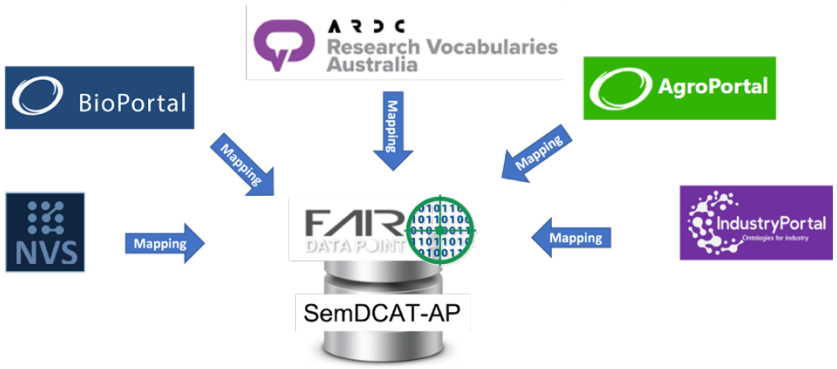

Figure 4. FAIRcat: harvest semantic artefact catalogues content and map them to a common profile.

Within FAIRcat, mappings link attributes present in the source catalogue to the equivalent ones in SemanticDCAT-AP. Those mappings are created in an editor depicted in Figure 5. The form is generated from the OWL specification, which allows populating all properties and rendering their name, definition, and level of endorsement (red "M" = mandatory). Because catalogues provide their metadata typically in JSON(-LD), the source metadata properties are identified using JSONPath.

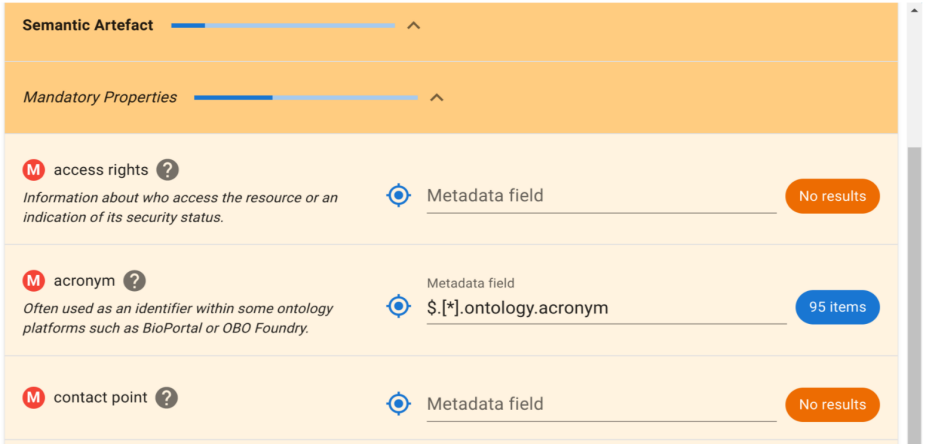

Figure 5. Illustration of mapping repository semantic artefacts to the SemanticDCAT-AP in FAIRcat.

Once the mapping step is finished, the harvested metadata can be converted into the SemanticDCAT-AP representation and stored in a FAIR Data Point [23] as illustrated in Figure 6. Apart from achieving SemanticDCAT-AP representation and FAIR-

[15] https://github.com/FAIRDataTeam/FAIRDataPoint

compliance, this approach enables harvesting multiple repositories into one FAIR Data Point that can be then searched; complex queries using SPARQL are also possible.

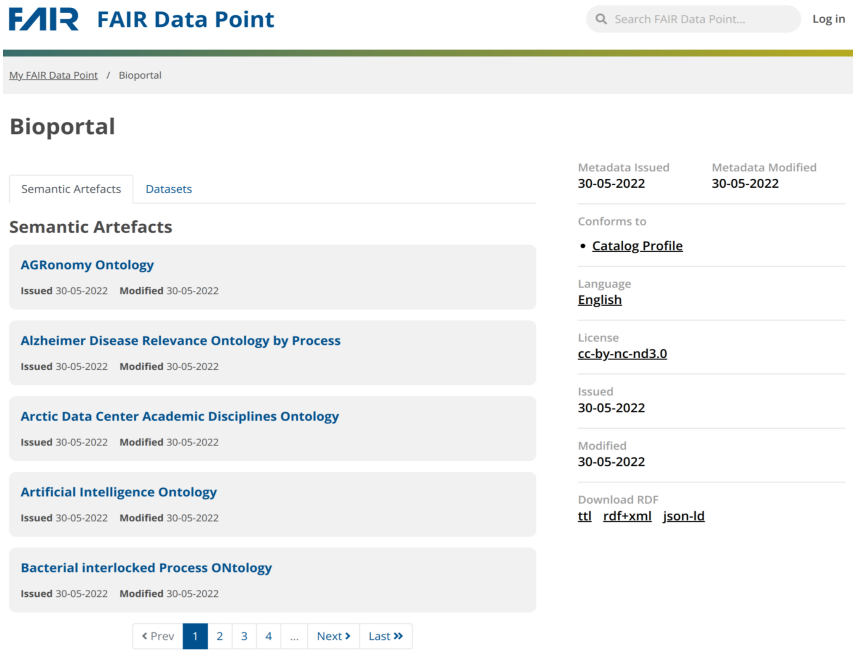

Figure 6. SemanticDCAT-AP metadata of the BioPortal semantic artefacts stored in a FAIR Data Point.

7. Conclusions and perspective

This paper presents the first attempt to define a common minimum metadata profile for FAIR semantic artefacts. This profile has been developed with the inputs from a large variety of communities. It aims to set a threshold, below which, an artefact can hardly be considered FAIR. Such a minimum metadata profile will be useful for FAIR assessment tools such as O'FAIRe and FOOPS! in their future evolutions.

With the FAIRCat prototype, three different semantic artefact catalogues have been harvested and mapped to this minimum metadata profile (via its implementation in SemanticDCAT-AP) in order to publish their content into a unique FAIR Data Point, allowing users to search across these three catalogues without copying the content or dealing with their specific APIs.

This work is now consolidated and refined in the context of FAIR-IMPACT i.e., the MOD2 proposition as well as the FAIRsFAIR profile (and its experimental implementation SemanticDCAT-AP). The aim is to reach a unified community-driven "standard to describe semantic artefacts. In the future, we also plan to investigate the W3C Profile Vocabulary (DX-PROF – https://www.w3.org/TR/dx-prof) to express the profile and eventually use it to provide a specification of a standard Application Programming Interface that semantic artefact catalogues could implement.

Finally, the SemanticDCAT-AP machine-actionable representations are an essential piece in implementing the FAIR principles as it can be used by the community to develop software tools, such as is the example of the FAIRcat prototype but also in FAIRness

assessment tools and any other relevant tools for semantic artefacts. Currently, FAIRcat is limited in its possibilities, for example it cannot harvest catalogues with complex APIs (or just in a limited way), but it demonstrates the idea. Currently, the mappings are stored into an in-house data model. As future work, these mappings will themselves be FAIRified, i.e., represented in a semantic way and stored with a persistent identifier.

Acknowledgments

We thank all the participants of the workshops and meetings organized by both the RDA VISSG and FAIRsFAIR project for time and fruitful discussions and feedback. This work has been originally supported in part by the H2020 FAIRsFAIR project (www.fairsfair.eu – grant #831558) as well as the *Data to Knowledge in Agronomy and Biodiversity* project (D2KAB – www.d2kab.org – ANR-18-CE23-0017). This work is now published –and continued– with support from the Horizon Europe FAIR-IMPACT project (https://fair-impact.eu – grant #101057344).

References

1. Batista, D., Gonzalez-Beltran, A., Sansone, S.-A., Rocca-Serra, P.: Machine actionable metadata models. Sci Data. 9, 592 (2022). https://doi.org/10.1038/s41597-022-01707-6.
2. Le Franc, Y., Coen, G., Essen, J.P., Bonino, L., Lehväslaiho, H., Staiger, C.: D2.2 FAIR Semantics: First recommendations. (2020). https://doi.org/10.5281/zenodo.3707985.
3. Hugo, W., Le Franc, Y., Coen, G., Parland-von Essen, J., Bonino, L.: D2.5 FAIR Semantics Recommendations Second Iteration. (2020). https://doi.org/10.5281/ZENODO.5362010.
4. Franc, Y. Le, Bonino, L., Koivula, H., Essen, J.P., Pergl, R.: D2.8 FAIR Semantics Recommendations Third Iteration. (2022). https://doi.org/10.5281/ZENODO.6675295.
5. Dutta, B., Toulet, A., Emonet, V., Jonquet, C.: New Generation Metadata vocabulary for Ontology Description and Publication. In: Garoufallou, E., Virkus, S., and Alemu, G. (eds.) 11th Metadata and Semantics Research Conference, MTSR'17. , Tallinn, Estonia (2017). https://doi.org/10.1007/978-3-319-70863-8_17.
6. Amdouni, E., Jonquet, C.: FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. In: Emmanouel Garoufallou and María-Antonia Ovalle-PerandonesAndreas Vlachidis (eds.) 15th International Conference on Metadata and Semantics Research, MTSR'21. pp. 67–80. Springer, Madrid, Spain (2021). https://doi.org/10.1007/978-3-030-98876-0_6.
7. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. Semantic Web and Information Systems. 5, 1–22 (2009). https://doi.org/10.4018/jswis.2009081901.
8. Janowicz, K., Hitzler, P., Adams, B., Kolas, D., Vardeman, C.: Five stars of Linked Data vocabulary use. Semant Web. 5, 173–176 (2014). https://doi.org/10.3233/SW-140135.
9. Hasnain, A., Rebholz-Schuhmann, D.: Assessing FAIR data principles against the 5-star open data principles. In: The Semantic Web: ESWC 2018 Satellite

Events. pp. 469–477. Springer, Heraklion, Greece (2018). https://doi.org/10.1007/978-3-319-98192-5_60/TABLES/1.
10. Garijo, D., Poveda-Villalón, M.: Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web. In: Cota, G., Daquino, M., and Pozzato, G.L. (eds.) Applications and Practices in Ontology Design, Extraction, and Reasoning. IOS Press (2020). https://doi.org/10.3233/SSW200034.
11. Matentzoglu, N., Malone, J., Mungall, C., Stevens, R.: MIRO: guidelines for minimum information for the reporting of an ontology. J Biomed Semantics. 9, 6 (2018). https://doi.org/10.1186/s13326-017-0172-7.
12. Jonquet, C., Toulet, A., Dutta, B., Emonet, V.: Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. Data Semantics. 7, 191–221 (2018). https://doi.org/10.1007/s13740-018-0091-5.
13. Garijo, D., Poveda-Villalón, M.: Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web. In: Cota, G., Daquino, M., and Pozzato, G.L. (eds.) Applications and Practices in Ontology Design, Extraction, and Reasoning. IOS Press (2020). https://doi.org/10.3233/SSW200034.
14. Poveda-Villalón, M., Espinoza-Arias, P., Garijo, D., Corcho, O.: Coming to Terms with FAIR Ontologies. In: Keet, M.C. and Dumontier, M. (eds.) 22nd International Conference on Knowledge Engineering and Knowledge Management, EKAW'20. pp. 255–270. Springer, Bolzano, Italy (2020). https://doi.org/10.1007/978-3-030-61244-3_18.
15. Garijo, D., Corcho, O., María Poveda-Villalón: FOOPS!: An Ontology Pitfall Scanner for the FAIR principles. In: 20th International Semantic Web Conference, ISWC'21: Posters, Demos, and Industry Tracks. CEUR Workshop Proceedings, 2980 (2021).
16. Coxid, S.J.D., Gonzalez-Beltranid, A.N., Magagna, B., Marinescu, M.-C.: Ten simple rules for making a vocabulary FAIR. (2021). https://doi.org/10.1371/journal.pcbi.1009041.
17. Zeng, M.L., Clunis, J.: FAIR + FIT: Guiding Principles and Functional Metrics for Linked Open Data (LOD) KOS Products. Journal of Data and Information Science. 5, 93–118 (2020). https://doi.org/10.2478/JDIS-2020-0008.
18. Frey, J., Streitmatter, D., Götz, F., Hellmann, S., Arndt, N.: DBpedia Archivo: A Web-Scale Interface for Ontology Archiving Under Consumer-Oriented Aspects. In: International Conference on Semantic Systems, SEMANTICS'20. pp. 19–35. Springer (2020). https://doi.org/10.1007/978-3-030-59833-4_2/TABLES/2.
19. Amdouni, E., Bouazzouni, S., Jonquet, C., O'faire, C.J.: O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources. Int J Metadata Semant Ontol. 16, 16–46 (2022). https://doi.org/10.13039/501100001665.
20. Amdouni, E., Bouazzouni, S., Jonquet, C.: O'FAIRe: Ontology FAIRness Evaluator in the AgroPortal Semantic Resource Repository. In: 19th Extended Semantic Web Conference, ESWC'22, Demo and Poster session. pp. 89–94. LNCS, Springer, Hersonissos, Crete, Greece (2022). https://doi.org/10.1007/978-3-031-11609-4_17.
21. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N.B., Jonquet, C., Rubin, D.L., Storey, M.-A., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res. 37, 170–173 (2009). https://doi.org/10.1093/nar/gkp440.

22. Côté, R.G., Jones, P., Apweiler, R., Hermjakob, H.: The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics. 7, 97 (2006). https://doi.org/10.1186/1471-2105-7-97.
23. Suarez-Figueroa, Hartmann, J., Sure, Y., Haase, P., Suarez-Figueroa, M.: OMV–ontology metadata vocabulary. In: Welty, C. (ed.) Workshop on Ontology Patterns for the Semantic Web, WOP'05. p. 9. Springer, Galway, Irland (2005).
24. Allocca, Carlo and d'Aquin, Mathieu and Motta, E.: DOOR - Towards a Formalization of Ontology Relations. In: International Conference on Knowledge Engineering and Ontology Development, KEOD'09. pp. 13–20. , Madera, Portugal (2009).