

# Towards a Conceptual Model for FAIR Metadata Schemas

Filipi Miranda Soares<sup>1,2,3</sup>, Luís Ferreira Pires<sup>3</sup>, Luiz Olavo Bonino da Silva Santos<sup>3,17</sup>, Rodrigo Fernandes Calhau<sup>3,4,5</sup>, Benildes Coura Moreira dos Santos Maculan<sup>6</sup>, Karen Coyle<sup>7</sup>, Shenghui Wang<sup>8</sup>, Erwin Folmer<sup>9</sup>, Debora Pignatari Drucker<sup>10</sup>, Maria Luiza de Almeida Campos<sup>11</sup>, Carlos Henrique Marcondes<sup>6,12</sup>, Maurício Barcellos Almeida<sup>6</sup>, Kelly Rosa Braghetto<sup>13</sup>, Guilherme Ataíde Dias<sup>14</sup>, José Augusto Salim<sup>15</sup>, Fernando Elias Corrêa<sup>2</sup>, Dilvan de Abreu Moreira<sup>2,16</sup>, Alexandre Cláudio Botazzo Delbem<sup>2,16</sup> and Antonio Mauro Saraiva<sup>1,2</sup>

<sup>1</sup>Polytechnic School, University of Sao Paulo, 158 Av. Prof. Luciano Gualberto, São Paulo - SP, 05508-010, Brazil

<sup>2</sup>Center for Artificial Intelligence (C4AI), 370 Av. Prof. Lúcio Martins Rodrigues, São Paulo - SP, 05508-020, Brazil

<sup>3</sup>Semantics, Cybersecurity & Services (SCS), Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, 5 Drienerlolaan, Enschede 7522 NB, The Netherlands

<sup>4</sup>LEDS, Federal Institute of Espírito Santo (IFES), 110 Av. Copacabana, Serra, 29173-087, Brazil

<sup>5</sup>Ontology and Conceptual Modeling Research Group, Federal University of Espírito Santo, 514 Av. Fernando Ferrari, Vitória - ES, 29075-910, Brazil

<sup>6</sup>Federal University of Minas Gerais, 6627 Av. Pres. Antônio Carlos, Pampulha, Belo Horizonte - MG, 31270-901, Brazil

<sup>7</sup>Independent Researcher, Berkeley, CA, USA

<sup>8</sup>Human Media Interaction (HMI), University of Twente, 5 Drienerlolaan, Enschede 7522 NB, The Netherlands

<sup>9</sup>HAN University of Applied Sciences, 26 Ruitenberglaan, Arnhem 6826 CC, The Netherlands

<sup>10</sup>Embrapa Agricultura Digital, 209 Av. Andre Tosello, Campus da Unicamp, Campinas - SP, 13083-886, Brazil

<sup>11</sup>Federal University of Bahia, 1946 Cláudio Fontenelle, Salvador - BA 40110-070, Brazil

<sup>12</sup>Programa de Pós-graduação em Ciência da Informação (PPGCI), Universidade Federal Fluminense, 126 R. Lara Vilela, Niterói - RJ, 24210-590, Brazil

<sup>13</sup>Institute of Mathematics and Statistics (IME), University of São Paulo, 1010 Rua do Matão, 05508-090 São Paulo - SP, Brazil

<sup>14</sup>Departamento de Ciência da Informação (CCSA), Universidade Federal da Paraíba, João Pessoa - PB, 58051-900, Brazil

<sup>15</sup>University of Campinas, 255 Monteiro Lobato Street, Campinas - SP 13083-862, Brazil

<sup>16</sup>Institute of Mathematics and Computer Sciences (ICMC), University of São Paulo, 400 Av. Trab. São Carlense, São Carlos - SP, 13566-590, Brazil

<sup>17</sup>BioSemantics, Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, Leiden 2333 ZC, The Netherlands

## Abstract

This paper explores the design and creation of metadata schemas based on the FAIR Data Principles. We provide a clear interpretation of these principles, focusing on how they apply to metadata schemas. Leveraging the OntoUML language, we developed a conceptual model that explains the key components of a FAIR-compliant metadata schema. Through detailed discussion and provision of examples for each model component, this work aims to help metadata designers and curators better understand how to incorporate the FAIR Data principles into their schemas.

## Keywords

Metadata schema, FAIR Data Principles, Conceptual Modeling, Semantic Web

# 1. Introduction

Metadata schemas are tools that provide a framework for creating consistent metadata records. Essentially, these schemas define data structures and represent the semantic model of specific domains. Ideally, their development should be community-driven to incorporate diverse insights and ensure broad applicability.

Over recent decades, many communities have developed a wide range of metadata schemas, significantly contributing to data standardization across domains. One of the goals of employing these schemas is to enhance data's findability, accessibility, interoperability, and reusability, in line with the FAIR principles. However, many metadata schemas do not adhere to the FAIR Data principles themselves, making their reuse difficult [1, 2, 3]. To solve this, schema curators should apply the FAIR principles in the development and management of these schemas. As outlined in FAIR principle I2, each community should decide the level of FAIRness required for the vocabularies they use. At a minimum, these vocabularies should have unique and permanent identifiers (F1), be accessible using standard methods (A1), and be described in a clear, widely understood language for knowledge representation (I1) [4].

This paper explores the application of FAIR Data principles to the design and management of metadata schemas. This research has been inspired by our work on agriculture and biodiversity data management. We propose an OntoUML conceptual model that outlines key components essential for metadata schemas to comply with these principles. The structure of the paper is as follows: Section 2 covers related work, Section 3 provides an introduction to OntoUML, Section

---

*ER2024: Companion Proceedings of the 43rd International Conference on Conceptual Modeling: ER Forum, Special Topics, Posters and Demos, October 28-31, 2024, Pittsburgh, Pennsylvania, USA*

✉ [f.mirandasoares@utwente.nl](mailto:f.mirandasoares@utwente.nl) (F. M. Soares); [lferreirapires@utwente.nl](mailto:lferreirapires@utwente.nl) (L. Ferreira Pires); [l.o.boninodasilvasantos@utwente.nl](mailto:l.o.boninodasilvasantos@utwente.nl) (L. O. Bonino da Silva Santos); [r.calhau@utwente.nl](mailto:r.calhau@utwente.nl) (R. F. Calhau); [benildes@gmail.com](mailto:benildes@gmail.com) (B. C. M. d. S. Maculan); [kcoyle@kcoyle.net](mailto:kcoyle@kcoyle.net) (K. Coyle); [shenghui.wang@utwente.nl](mailto:shenghui.wang@utwente.nl) (S. Wang); [Erwin.folmer@han.nl](mailto:Erwin.folmer@han.nl) (E. Folmer); [deboradrucker@embrapa.br](mailto:deboradrucker@embrapa.br) (D. P. Drucker); [marialuizalmeida@gmail.com](mailto:marialuizalmeida@gmail.com) (M. L. d. A. Campos); [ch\\_marcondes@id.uff.br](mailto:ch_marcondes@id.uff.br) (C. H. Marcondes); [mba@eci.ufmg.br](mailto:mba@eci.ufmg.br) (M. B. Almeida); [kellyrb@ime.usp.br](mailto:kellyrb@ime.usp.br) (K. R. Braghetto); [guilhermeataide@gmail.com](mailto:guilhermeataide@gmail.com) (G. A. Dias); [jasalim@unicamp.br](mailto:jasalim@unicamp.br) (J. A. Salim); [fecorrea@usp.br](mailto:fecorrea@usp.br) (F. E. Corrêa); [dilvan@icmc.usp.br](mailto:dilvan@icmc.usp.br) (D. d. A. Moreira); [acbd@icmc.usp.br](mailto:acbd@icmc.usp.br) (A. C. B. Delbem); [saraiva@usp.br](mailto:saraiva@usp.br) (A. M. Saraiva)

🌐 <https://research.utwente.nl/en/persons/filipi-miranda-soares> (F. M. Soares); <https://research.utwente.nl/en/persons/luis-ferreira-pires> (L. Ferreira Pires); <https://people.utwente.nl/l.o.boninodasilvasantos> (L. O. Bonino da Silva Santos); <https://nemo.inf.ufes.br/equipe/rodrigo-calhau/> (R. F. Calhau); <http://kcoyle.net> (K. Coyle); <https://people.utwente.nl/shenghui.wang> (S. Wang); <https://www.han.nl/onderzoek/onderzoekers/erwin-folmer/> (E. Folmer); <https://www.embrapa.br/equipe/-/empregado/353696/deboradrucker> (D. P. Drucker); <http://lattes.cnpq.br/9545682339961651> (M. L. d. A. Campos); <https://mba.eci.ufmg.br/english/> (M. B. Almeida); <http://lattes.cnpq.br/7588616253679294> (J. A. Salim)

🆔 0000-0002-0674-7960 (F. M. Soares); 0000-0001-7432-7653 (L. Ferreira Pires); 0000-0002-1164-1351 (L. O. Bonino da Silva Santos); 0009-0006-6051-2165 (R. F. Calhau); 0000-0003-4303-9071 (B. C. M. d. S. Maculan); 0000-0002-1085-7089 (K. Coyle); 0000-0003-0583-6969 (S. Wang); 0000-0002-7845-1763 (E. Folmer); 0000-0003-4177-1322 (D. P. Drucker); 0000-0002-9253-3706 (M. L. d. A. Campos); 0000-0003-0929-8475 (C. H. Marcondes); 0000-0002-4711-270X (M. B. Almeida); 0000-0001-6218-6849 (K. R. Braghetto); 0000-0001-6576-0017 (G. A. Dias); 0000-0002-8675-7068 (J. A. Salim); 0000-0002-2744-6104 (F. E. Corrêa); 0000-0002-4801-2225 (D. d. A. Moreira); 0000-0003-1810-1742 (A. C. B. Delbem); 0000-0003-2283-1123 (A. M. Saraiva)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

4 outlines the methods, Section 5 presents a conceptual model for FAIR metadata schemas, Section 6 focuses on assessing the FAIRness of metadata schemas, and Section 7 concludes with final considerations.

## 2. Related Work

The application of FAIR data principles to semantic artifact curation is an evolving field, crucial for ensuring that resources like controlled vocabularies, taxonomies, ontologies, and metadata schemas support data FAIRification. Several studies have explored this area.

Berg-Cross and Arbor [5] highlight the importance of ontologies in making metadata computer-processable by providing axiomatized definitions. However, adapting FAIR principles to these resources introduces challenges, such as the need for community standards and meaningful semantics.

The FAIRsFAIR project report [1] proposes refined principles to improve the FAIRness of semantic artifacts, building on existing community recommendations. Poveda-Villalón et al. [3] examine the challenges of aligning ontologies with FAIR principles, emphasizing the need for ontologies that support data interoperability and reuse. Amdouni and Jonquet [2] present a framework for evaluating the FAIRness of semantic artifacts, offering a quantitative approach to assess adherence to FAIR principles. These efforts underscore the importance of establishing robust standards and practices for FAIR semantic artifacts construction, however none of them target metadata schemas specifically, neither present a conceptual model to elucidate the requirements for FAIR metadata schemas. Considering the importance of community metadata schemas for FAIR data, this paper proposes a series of recommendations to make this specific type of semantic artifact FAIR.

## 3. A (very) Brief Introduction to OntoUML Stereotypes

OntoUML is a language for specifying ontologies that has been implemented as a UML extension. Grounded in the Unified Foundational Ontology (UFO), OntoUML provides a robust conceptual framework, making it an ideal tool for accurately representing complex domains [6, 7], such as in this paper. UFO's comprehensive approach to ontology modeling, which includes key elements of real-world phenomena such as objects, events, and their interrelations, forms a solid basis for OntoUML. Consequently, OntoUML leverages this foundational strength to create "detailed and precise" syntactic models; however, semantics remain within the model users' agreement. [6, 7].

OntoUML introduces several stereotypes to capture different kinds of ontological distinctions. These stereotypes offer a clear and precise way to model various elements within a domain, including, but not limited to: `Kind`, a natural category that provides identity to its instances; `Role`, an entity's role in a specific context; `Subkind`, a specialized kind that inherits properties from a broader kind while introducing unique features; `Mixin`, a characteristic shared by entities from different kinds; `RoleMixin`, which combines the concepts of role and mixin to represent a role that can be assumed by entities from different kinds; `Category`, an abstract grouping of mixins that does not provide identity; `Relator`, which mediates relationships between two or

more entities; and *Quality*, a property or characteristic that an entity possesses, which can be measured or perceived [6].

OntoUML is particularly useful in scenarios where precise and unambiguous domain representations are crucial. For instance, in the development of FAIR metadata schemas, OntoUML can help create models that allows us to understand all the components and their relationships to define a FAIR metadata schema.

## 4. Methods

We defined our ontological conceptual model by following an approach similar to that proposed by Guizzardi [8], known as “ontological unpacking”. This method involves analyzing a domain to reveal its underlying ontological conceptual model. Collaboration with specialists provided valuable insights that shaped this model. The following steps were undertaken:

**Domain Analysis:** This phase focused on identifying key concepts, entities, and relationships within the domain through structured meetings, workshops, and communications. Key discussions occurred during workshops promoted by the GO FAIR Agro Brazil Network<sup>1</sup>, Brazil’s 5th National Action Plan on Open Government<sup>2</sup>, and the development of the Almes Core Metadata Schema [9]. Additionally, collaboration with the TDWG Biological Interaction Data Interest Group<sup>3</sup> during the development of the Plant-Pollinator Interactions (PPI) Vocabulary [10] played a crucial role. Insights from the WorldFAIR Case Study on Agricultural Biodiversity [11, 12] also helped us to understand the key components of FAIR community metadata schemas.

**Literature Consultation:** Literature suggestions from domain experts and related works were also considered. The work of Marcia L. Zeng and collaborators [13, 14], and the ISO/IEC 11179 standards [15, 16, 17, 18] were key references for several concepts included in this model, due to their solid foundation for metadata schema modeling. The domain analysis combined with insights from the literature resulted in the requirements listed in Table 1.

**Ontological Analysis:** We examined the concepts in Table 1 to determine their ontological nature, and then classified these concepts into appropriate ontological categories, such as kinds, roles, phases, and mixins, in accordance with OntoUML.

**Model Development:** The conceptual model was developed based on the insights gained from the ontological analysis. This model represents the ontological structure of the domain of FAIR metadata schemas, capturing the essential nature and relationships of its concepts.

**Validation:** We validated the model by conducting a FAIR assessment grounded in the concepts outlined in this conceptual model.

---

<sup>1</sup><https://go-fair-agro.github.io/>

<sup>2</sup>[https://www.opengovpartnership.org/wp-content/uploads/2022/01/Brazil\\_Action-Plan\\_2021-2023\\_EN.pdf](https://www.opengovpartnership.org/wp-content/uploads/2022/01/Brazil_Action-Plan_2021-2023_EN.pdf)

<sup>3</sup><https://github.com/tdwg/interaction>

<sup>4</sup><https://www.dcc.ac.uk/guidance/standards/metadata>

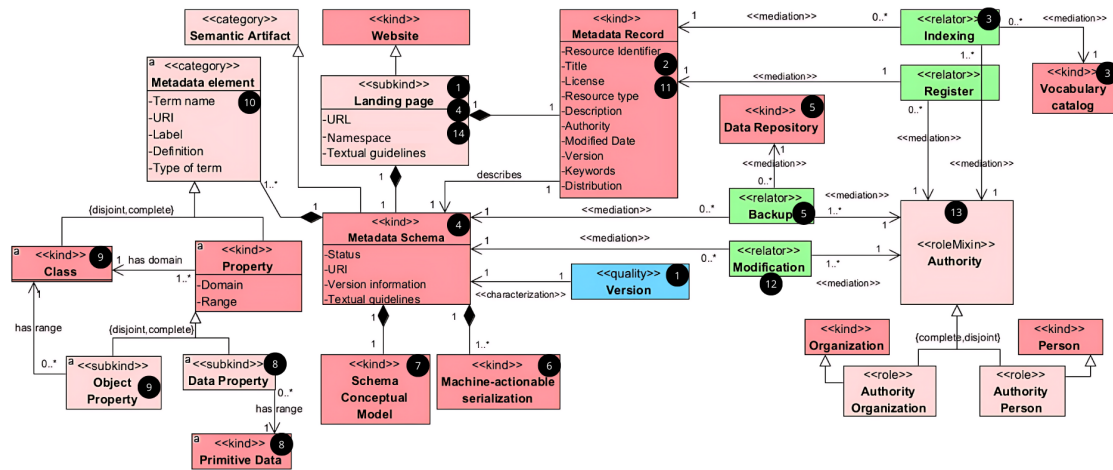
**Table 1**

Adaptation of the FAIR Data Principles [4, 19] based on [1, 13] for metadata schema management.

FAIR P.	Ref. P.	Refinement explanation
F1	RF1	Each metadata schema should have a unique identifier (i.e., the schema namespace), represented by a Uniform Resource Identifier (URI). Different versions of a metadata schema should have distinct identifiers.
F2	RF2	The landing page of a metadata schema should contain a descriptive metadata record with details about the schema. This metadata record should be machine-readable to allow search engines and other automated tools to discover the schema [20].
F4	RF3	To improve metadata schema findability, the schema and its metadata record should be indexed in vocabulary catalogs such as the Linked Open Vocabularies (LOV) [21], OntoBee [22], and the Digital Curation Centre List of Metadata Standards <sup>4</sup>
A1, A1.1	RF4	The URIs implemented by the metadata schema should resolve to web pages accessible through widely used, universally accessible and secure communication protocols, such as HTTPS.
A2	RF5	The sustained accessibility of metadata schema documentation should be ensured, regardless of whether the schema is still active or maintained. This is important for applications that use past schema versions.
I1	RF6	Formal languages (e.g., RDF, OWL) should be used for the machine-readable representation of metadata schemas.
I1	RF7	Conceptual models in UML, OntoUML, or other languages should be provided to help humans understand the relationships between concepts.
I2	RF8	The range (data types) and the domain (a class) of datatype properties should be clearly specified in the schema.
I2, R1	RF9	The range (a class) and the domain (a class) for object properties should be clearly specified. If the classes used as range are defined elsewhere, this external controlled vocabulary should also adhere to the FAIR principles.
I2	RF10	Properties and classes should have clear definitions and names.
R1.1	RF11	Metadata schemas should be published under open licenses such as CC BY 4.0, and the license information should be machine-readable.
R1.2	RF12	Schema modifications, such as reviewing, deleting, and adding new terms should be documented.
R1.2	RF13	The authority responsible for curating the metadata schema should be indicated.
R1.3, I3	RF14	When developing new metadata schemas, reusing elements from existing schemas should be prioritized to minimize duplication. Reuse is recommended only when the elements have the same meaning, regardless of the schema context. The namespaces of all reused terms should be listed on the schema's landing page and included in its serializations.

## 5. A Conceptual Model for FAIR Metadata Schemas

Drawing on the refined FAIR Data principles in Table 1, supplemented by insights from the literature and the collective domain expertise of this paper's collaborators, we present a conceptual model for FAIR metadata schemas in Fig. 1. We further explain each component of the conceptual model, providing examples.



**Figure 1:** Conceptual Model of a FAIR Metadata Schema. Black circles represent refined FAIR principles from Table 1.

### 5.1. Schema Authority

A FAIR metadata schema should be created through the collaborative efforts of specific communities of practice, to ensure they encode their discipline-specific elements [23]. Examples of community-developed schemas include DwC [24], Astronomy Visualization Metadata (AVM) [25], and the Data Documentation Initiative (DDI) [26]. Expert groups or institutions, often referred to as the schema authority [15, 17, 18, 27], should lead the development and curation of metadata schemas. The authority is responsible for any modification in the schema, which includes adding, reviewing, and removing metadata elements.

The authority behind the metadata schema should be publicly declared on the schema’s landing page and within the schema metadata record. This transparency ensures that the user community knows who to contact for assistance, inquiries, suggestions, or participation in the curation community. The authority is also responsible for indexing the schema (and its descriptive metadata record) in vocabulary catalogs like Linked Open Vocabularies (LOV) and OntoBee, so more users can find and reuse it.

Additionally, maintaining a backup of the metadata schema in data repositories like GitHub and Zenodo is crucial for its longevity, accessibility, and version control. Platforms like GitHub provide tools for tracking changes, reverting to previous versions, and facilitating effective collaboration.

When setting up a backup on GitHub, including a README file that explains the schema’s purpose and provides guidelines for use and contribution is helpful. Utilizing GitHub features such as branches for ongoing developments, tags for stable releases, and releases for distributing specific versions can improve schema management. Engaging the community through issues and pull requests can also gather feedback and contributions, enriching the schema’s development and adoption. Examples of community schemas using GitHub for this purpose include DwC<sup>5</sup>,

<sup>5</sup><https://github.com/tdwg/dwc>



DDI-CDI<sup>6</sup>, and Almes Core (ALM)<sup>7</sup> [9].

## 5.2. Landing Page of a Metadata Schema

The landing page of a metadata schema is a type of website that consolidates all the information describing a schema. It typically includes details such as the schema's purpose, structure, and usage guidelines. It should be accessible via a URL that supports open data transfer protocols such as HTTPS, as recommended by RF5. This page should include at least the following components:

- A schema metadata record, describing the schema in formats accessible to both humans and machines.
- The metadata schema itself, providing the structure and specifications.

Furthermore, the landing page should declare the namespaces employed by the metadata schema, ensuring consistent terms usage. This includes declaring namespaces for terms reused from other schemas as well. For instance, The Data Catalog Vocabulary (DCAT) defines its namespace as <http://www.w3.org/ns/dcat#>, reuses many terms from Dublin Core (DC), and introduces a minimal set of its own classes and properties<sup>8</sup>.

## 5.3. Schema's Metadata Record

A schema metadata record typically appears in the first section of the landing page, as in the W3C standards, for instance. It should be accessible and readable to both humans and machines. This record should include, at least: resource identifier (the schema URI), title, distribution, license, resource type, description, authority, modified date, version, keywords, and distributions. This aligns with the recommendations for metadata discovery by the Cross Domain Interoperability Framework (CDIF) [20], which advises on using metadata terms from established schemas like DCAT, Schema.org (SDO), and DC to create the metadata record, and recommends using machine readable formats like JSON-LD. The metadata record should be provided (or registered) by the schema's authority.

## 5.4. Metadata Schema

A metadata schema is a type of semantic artifact aimed at organizing, categorizing, and managing information and knowledge within a specific domain or context [28, 29]. A metadata schema requires comprehensive documentation to ensure it is accessible and reusable for humans and computer applications alike. This documentation comprehends several components, as detailed in Fig. 1. A key component of metadata schemas is its namespace, which distinguishes it from other schemas.

Metadata schemas have as basic components metadata elements and the relationships between these elements [13]. In an ontological approach, these elements are usually classified

---

<sup>6</sup><https://github.com/ddialliance>

<sup>7</sup><https://github.com/AlmesCore/>

<sup>8</sup><https://www.w3.org/TR/2023/WD-vocab-dcat-3-20230307/>

as classes or properties [13, 30]. They are typically described in textual form (as a glossary of terms) on the schema’s landing page for human consumption. Additionally, these schemas should be represented in both conceptual and machine-actionable models.

## 5.5. Metadata Element

A metadata element can be either a class or a property [13]. A class represents a category or type of thing (e.g., “Book,” “Person,” “Organism”) and is a way to group similar items based on shared characteristics [31]. Classes provide a high-level organizational structure to metadata. They allow for categorizing different entities or concepts within the described domain. For instance, the DwC metadata schema (developed for the biodiversity domain) includes classes such as `dwc:Event`, `dwc:Occurrence`, and `dwc:Taxon`.

Properties describe specific characteristics or aspects of an individual within a given class. For the class `dwc:Event`, examples of properties include `dwc:eventDate`, `dwc:eventID`, and `dwc:EventMeasurement`. Properties in metadata schemas, as in ontologies, can be classified into two types: object property and data property.

Object properties define relationships between two entities (or classes) within the metadata schema. They link instances of one class to instances of another class [32]. For example, in a bibliographic metadata schema:

```
:author a owl:ObjectProperty ;  
    rdfs:domain :Book ;  
    rdfs:range :Person .
```

In this example, `:author` is an object property that relates a `:Book` to a `:Person`. This indicates that a book can have one or more authors, and each author is a person.

Data properties (also known as datatype properties) define relationships between an entity and primitive data types. They link instances of a class to values such as strings, numbers, or dates (i.e., primitive data types) [32, 33]. For example, in the bibliographic schema:

```
:publicationDate a owl:DatatypeProperty ;  
    rdfs:domain :Book ;  
    rdfs:range xsd:date .
```

The domain of a property specifies the class of individuals to which the property can be applied. In other words, if a property has a certain class as its domain, then any individual that possesses this property is assumed to be a member of that class. For instance, if a property `:hasAuthor` has a domain `:Book`, then any resource that `:hasAuthor` is inferred to be a `Book` [32].

Defining domains and ranges between concepts in a metadata schema enhances the precision and clarity of the relationships between data elements and enables machine inferences [34]. By specifying which classes of objects a property can relate to (domain) and the type of values it can take (range), we infer that data elements belong to the specified classes.

In addition to the relationships between elements in a schema, each metadata element should include a specification of semantics and content [13, 14, 15]. Semantics is about the “meanings of the elements”, and content is related to providing “general instructions on what and how values should be assigned to the elements in an application” [13, p.39].



## 5.6. Element Semantic Specification

Unclear semantic specification of the elements in a metadata schema hamper its reuse. A combination of attributes can be used to define the semantics of an element. For instance in DC, “term name” assigns a specific, distinct name to the metadata element within the schema. A “Label” is included for human readability and can be available in multiple languages. “Definition” provides in-depth explanations regarding the element’s purpose, scope and use.

The ISO/IEC 11179 standards, grounded on the Theory of Terminology, guide articulating the semantics of metadata elements, focusing on formulating definitions in Part 4 [17] and term names in Part 5 [18].

### 5.6.1. Formulating Metadata Definitions

Part 4 of ISO/IEC 11179 outlines a set of requirements and recommendations for metadata definitions [17]. The requirements specify that definitions should focus on singular concepts, express the essence of the concept rather than what it is not, and be formulated as descriptive phrases or sentences. They should also avoid uncommon abbreviations and not embed definitions of other data or underlying concepts. Additionally, the recommendations advise that definitions convey the essential meaning of the concept, be precise, unambiguous, concise, and able to stand alone. Definitions should also avoid embedding rationale, functional usage, domain information, or procedural details. Circular reasoning should be avoided, and consistent terminology and logical structure should be maintained for related definitions.

### 5.6.2. Formulating Metadata Names

In addition to providing clear definitions, metadata elements must have appropriate names. ISO/IEC 11179-5 [18] provides naming conventions to ensure names are meaningful and easily understood by users. In this standard, each name represents a concept in a metadata schema and follows specific conventions, which include syntactic, semantic, and lexical rules that should be documented. The guiding principles for naming include defining the scope (general or specific), identifying the naming authority, specifying the components (semantic principle), arranging them (syntactic principle), and determining their appearance (lexical principle). Additionally, the uniqueness principle ensures names are distinct within the defined scope, with differentiation rules if needed.

ISO/IEC 11179-5 provides a flexible framework for nomenclature, adaptable across domains. The semantics of name parts depend on the type of metadata term. While class-type terms may be abstract, they must still be named and hierarchically organized within the schema, similar to other metadata terms, in relation to their properties [18].

Property name consistency follows two main rules: **Discreteness**, where each property term has a unique meaning without overlap, and **Completeness**, ensuring that the set of property terms fully describes an entity [18]. Property names can indicate value domains, such as Color, Quantity, or Height, as seen in the Plant-Pollinator Interactions (PPI) schema, with terms like `ppi:styleLengthInMilimeters`, `ppi:ovuleQuantity`, and `ppi:visitedFlowersQuantity` [10].

A property name may also include a qualifier term [18]. A qualifier can be used to distinguish metadata terms with identical names or to represent more specific values, such as the qualifiers ‘start’ and ‘end’ in `dcat:startDate` and `dcat:endDate`.

Finally, naming conventions should include the use of separators like hyphens, grouping symbols, or formatting resources like CamelCase, to connect name parts, for both classes and properties [18].

## 5.7. Machine-Actionable Serialization

Semantic Web languages like RDF and OWL allow the creation of machine-readable and actionable serializations of metadata schemas. Initially, early metadata schemas, including DC and VRA Core, were designed as purely XML schemas. Over time, particularly in the second decade of the millennium, there was a shift towards RDF vocabularies [13].

FAIR metadata schemas should be implemented as Semantic Web schemas, using languages such as RDF and OWL, as outlined in principle RF7 in Table 1. These languages enable advanced functions like semantic querying and automated reasoning. RDF provides a framework for expressing information about resources in a graph form, which is highly beneficial for linking data [35]. OWL adds a richer ontology layer, representing more complex relationships and classifications, enhancing the depth and utility of the schema [36]. The choice of serialization language depends on the goals governing authority of the schema. A persistent identifier (URI or IRI) should be used as the schema namespace. Each metadata element should then have a unique identifier within that namespace. There are services that help generate these identifiers, such as the W3C Permanent Identifier Community Group<sup>9</sup> and `purl.org`<sup>10</sup>.

## 5.8. Schema Conceptual Model

A conceptual model is pivotal in developing and implementing a metadata schema. It is an abstract representation of the phenomena the metadata schema aims to describe, capturing the essential concepts and their relationships, sometimes referred to as metamodel or upper model. A conceptual model may serve different purposes in a metadata schema:

- Schema design: A conceptual model helps identify the key concepts, entities, and attributes relevant to the described domain, ensuring that the schema covers all necessary aspects of the data it intends to represent [37].
- Data integration and mapping: Conceptual models may provide a reference point for mapping between schemas. They are a fundamental component for enabling semantic interoperability between heterogeneous datasets [37].
- Human understanding: Conceptual models are useful for schema designers and end-users. They can facilitate the understanding of the schema by providing a visual, conceptual representation of its classes and properties. This also makes the schema easier to reuse by users unfamiliar with it [37].

---

<sup>9</sup><https://w3id.org/>

<sup>10</sup><https://purl.org/>

Many metadata schemas use conceptual models for these purposes. UML is one of the most popular language used for metadata schema conceptual models. For instance, EML<sup>11</sup> is based on a UML conceptual model that defines the structure and semantics of metadata to describe ecological data effectively. Besides EML, the DDI-CDI Framework is another example of a schema that also uses a UML conceptual model, which is described as the “core” of their schema<sup>12</sup>.

## 6. Assessing the FAIRness of Metadata Schemas

Assessing the FAIRness of digital objects is a complex, multi-layered task that involves evaluating various elements contributing to compliance with the FAIR principles. This process can be either automated or manual, each with its own set of advantages and limitations. Automated assessments typically rely on widely accepted metadata schemas (e.g., DC, SDO, DCAT) and identifier systems such as PURL and DOI. While these tools are efficient, they may overlook domain-specific schemas or resources that do not fit within their predefined rules, potentially resulting in lower FAIRness scores [38]. Manual assessments offer more flexibility to evaluate domain-specific elements, but this approach can introduce bias due to subjective interpretation [38].

In this paper we introduce a methodology for manually assessing the FAIRness of metadata schemas, applying principles similar to those in Amdouni and Jonquet’s FAIRness Assessment Grid for Ontologies [2]. The proposed approach attributes scores to FAIR indicators, which are then summed up to calculate an overall FAIR compliance percentage for each principle.

Using this method, we assessed the FAIRness of three metadata schemas: Darwin Core (DwC)<sup>13</sup>, D-CAT<sup>14</sup>, and Almes Core (ALM)<sup>15</sup>. The overall FAIR compliance scores were as follows: DwC: approximately 87%, D-CAT: 100%, and ALM: 91.75%. The main reason DwC did not achieve 100% was the lack of machine-actionable serializations and defined domains and ranges between classes and properties. ALM’s score for findability was reduced because it is not indexed in vocabulary catalogs. Detailed guidelines and examples of how this assessment was applied to these specific metadata schemas are provided on Zenodo [1].

These assessments illustrate the applicability of the proposed conceptual model and demonstrate its utility in helping communities identify and address gaps in their metadata schemas to achieve full FAIR compliance.

## 7. Final Considerations

The FAIR assessment proposed here is not intended as an indicator of schema quality. As emphasized by GO FAIR [4] and mentioned in this paper’s introduction, the responsibility for determining the FAIRness level of schemas rests with the goals set by the schema authorities.

---

<sup>11</sup><https://eml.ecoinformatics.org>

<sup>12</sup>[https://ddialliance.org/Specification/ddi-cdi#uml\\_model](https://ddialliance.org/Specification/ddi-cdi#uml_model)

<sup>13</sup><http://rs.tdwg.org/dwc/terms/>

<sup>14</sup><http://www.w3.org/ns/dcat#>

<sup>15</sup><https://w3id.org/AlmesCore#>

This paper aims to support communities in applying FAIR Data principles to metadata schema management, offering insights to enhance the findability, accessibility, interoperability, and reusability of their schemas.

Regarding limitations, the proposed FAIR assessment requires further refinement. Developing an automated tool to streamline the FAIR assessment process would be a valuable enhancement. For future work, additional empirical validation through multiple case studies or user feedback from metadata schema designers is necessary to further mature and validate this conceptual model.

## Data Availability

1. Soares, F. M., Ferreira Pires, L., Bonino da Silva Santos, L. O., Calhau, R. F., Coura Moreira dos Santos Maculan, B., Coyle, K., Wang, S., Folmer, E., Drucker, D., de Almeida Campos, M. L., Marcondes, C. H., Barcellos Almeida, M., Braghetto, K. R., Ataíde Dias, G., Salim, J. A., Correa, F. E., Botazzo Delbem, A. C., & Saraiva, A. M. (2024). Accessing the FAIRness of Metadata Schemas. Zenodo. <https://doi.org/10.5281/zenodo.13830523>

## Acknowledgments

FMS received funding from the São Paulo Research Foundation - FAPESP (Grant n. 2021/15125-0 and 2022/08385-8). BCMSM received funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Grant n. 307765/2023-7). KRB received funding from FAPESP (Grant n. 2023/00779-0). FMS, DPD, KRB, FEC, DAM, ACBD, and AMS thank the Center for Artificial Intelligence (C4AI), a partnership of USP, IBM, and FAPESP (Grant 2019/07665-4). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001.

## References

- [1] Y. Le Franc, J. Parland-von Essen, L. Bonino, et al., D2.2 fair semantics: First recommendations, 2020. doi:10.5281/ZENODO.5361930.
- [2] E. Amdouni, C. Jonquet, Fair or fairer? an integrated quantitative fairness assessment grid for semantic resources and ontologies, in: E. Garoufallou, M. A. Ovalle-Perandones, A. Vlachidis (Eds.), *Metadata and Semantic Research. MTSR 2021*, volume 1537 of *Communications in Computer and Information Science*, Springer, Cham, 2022. doi:10.1007/978-3-030-98876-0\_6.
- [3] M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, O. Corcho, Coming to terms with fair ontologies, in: C. M. Keet, M. Dumontier (Eds.), *Knowledge Engineering and Knowledge Management. EKAW 2020*, volume 12387 of *Lecture Notes in Computer Science*, Springer, Cham, 2020. doi:10.1007/978-3-030-61244-3\_18.
- [4] GO FAIR, Fair principles, <https://www.go-fair.org/fair-principles/>, 2023. Accessed: 2024-08-02.

- [5] G. Berg-Cross, S. Arbor, Beyond simple fair principles for ontologies and semantic resources: Grounding rich, meaningful metadata, *Journal of the Washington Academy of Sciences* 108 (2022) 1–26. URL: <https://www.jstor.org/stable/27281297>, accessed 2 Aug. 2024.
- [6] G. Guizzardi, *Ontological Foundations for Structural Conceptual Models*, Ph.D. thesis, University of Twente, 2005.
- [7] G. Guizzardi, G. Wagner, Conceptual simulation modeling with onto-uml advanced tutorial, in: *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 2012, pp. 1–15. doi:10.1109/WSC.2012.6465133.
- [8] G. Guizzardi, A. Bernasconi, O. Pastor, V. C. Storey, Ontological unpacking as explanation: The case of the viral conceptual model, in: A. Ghose, J. Horkoff, V. E. Silva Souza, J. Parsons, J. Evermann (Eds.), *Conceptual Modeling*, Springer International Publishing, Cham, 2021, pp. 356–366.
- [9] F. M. Soares, F. E. Corrêa, L. F. Pires, L. O. B. da Silva Santos, D. P. Drucker, et al., Building a community-based fair metadata schema for brazilian agriculture and livestock trading data, in: *Proceedings of the SEMPW 2022 Workshop*, CEUR Workshop Proceedings, Rheinisch-Westfälische Technische Hochschule, 2022.
- [10] J. A. Salim, A. M. Saraiva, P. F. Zermoglio, K. Agostini, M. Wolowski, et al., Data standardization of plant–pollinator interactions, *GigaScience* 11 (2022) giac043. doi:10.1093/gigascience/giac043.
- [11] D. Drucker, J. Salim, J. Poelen, et al., *WorldFAIR (D10.2) Agricultural Biodiversity Standards, Best Practices and Guidelines Recommendations*, Zenodo, 2024. doi:10.5281/ZENODO.10666593.
- [12] D. Drucker, J. Salim, J. Poelen, et al., *WorldFAIR (D10.3) Agricultural biodiversity FAIR data assessment rubrics*, Zenodo, 2024. doi:10.5281/zenodo.10719265.
- [13] M. L. Zeng, J. Qin, *Metadata*, third edition ed., ALA Neal-Schuman, Chicago, 2022.
- [14] L. M. Chan, M. L. Zeng, *Metadata interoperability and standardization - a study of methodology part i: Achieving interoperability at the schema level*, *D-Lib Magazine* 12 (2006). doi:10.1045/june2006-chan.
- [15] ISO/IEC, *ISO/IEC 11179-1:2023: Information technology: Metadata registries (MDR): Part 1: Framework*, 2023.
- [16] ISO, *Building a Metadata Schema – Where to Start*, ISO, Geneva, 2008.
- [17] ISO/IEC, *Information technology – metadata registries (mdr) – part 4: Formulation of data definitions*, 2004.
- [18] ISO/IEC, *Information technology – metadata registries (mdr) – part 5: Naming principles*, 2015.
- [19] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, et al., *The fair guiding principles for scientific data management and stewardship*, *Scientific Data* 3 (2016) 160018. doi:10.1038/sdata.2016.18.
- [20] *Cross-Domain Interoperability Framework (CDIF) Working Group*, S. Richard, A. Gregory, S. Hodson, D. Fils, et al., *Cross domain interoperability framework (cdif): Discovery module (v01 draft for public consultation)*, 2023. doi:10.5281/ZENODO.10252564.
- [21] P.-Y. Vandenbussche, G. A. Ateazing, M. Poveda-Villalón, B. Vatant, *Linked open vocabularies (lov): A gateway to reusable semantic vocabularies on the web*, *Semantic Web* 8

- (2017) 437–452.
- [22] E. Ong, Z. Xiang, B. Zhao, et al., Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration, *Nucleic Acids Research* 45 (2017) D347–D352. doi:10.1093/nar/gkw918.
  - [23] M. A. Musen, M. J. O’Connor, E. Schultes, et al., Modeling community standards for metadata as templates makes data fair, *Scientific Data* 9 (2022) 696. doi:10.1038/s41597-022-01815-3.
  - [24] J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, et al., Darwin core: An evolving community-developed biodiversity data standard, *PLoS ONE* 7 (2012) e29715. doi:10.1371/journal.pone.0029715.
  - [25] R. L. Hurt, A. Gauthier, L. Christensen, R. Wyatt, Sharing images intelligently: The astronomy visualization metadata standard, in: *CAP’07: Proceedings of the Conference on Communicating Astronomy with the Public*. Eugenides Foundation, Citeseer, 2007, pp. 450–453.
  - [26] M. Vardigan, P. Heus, W. Thomas, Data documentation initiative: Toward a standard for the social sciences, *International Journal of Digital Curation* 3 (2008).
  - [27] S. Higgins, What are metadata standards, <https://www.dcc.ac.uk/guidance/briefing-papers/standards-watch-papers/what-are-metadata-standards>, 2007. Accessed: 04/12/2023.
  - [28] F. Mazzocchi, Knowledge organization system (kos): An introductory critical account, *KNOWLEDGE ORGANIZATION* 45 (2018) 54–78. doi:10.5771/0943-7444-2018-1-54.
  - [29] G. M. Hodge, Systems of knowledge organization for digital libraries: beyond traditional authority files, volume 91, Digital Library Federation, 2000.
  - [30] J. Riley, *Understanding Metadata*, 4th ed., National Information Standards Organization, Bethesda, MD, USA, 2017.
  - [31] J. M. Reitz, Online dictionary for library and information science, 2023. URL: [https://odlis.abc-clio.com/odlis\\_c.html](https://odlis.abc-clio.com/odlis_c.html).
  - [32] D. Brickley, R. Guha, Rdf schema 1.1, 2014. URL: <http://www.w3.org/TR/rdf11-schema/>.
  - [33] P. Panov, L. N. Soldatova, S. Džeroski, Generic ontology of datatypes, *Information Sciences* 329 (2016) 900–920. doi:10.1016/j.ins.2015.08.006, special issue on Discovery Science.
  - [34] Dublin Core Metadata Initiative, Domains and ranges, 2024. URL: [https://www.dublincore.org/resources/glossary/domains\\_and\\_ranges/](https://www.dublincore.org/resources/glossary/domains_and_ranges/), accessed: 2024-05-28.
  - [35] RDF Working Group, Resource description framework (rdf), <https://www.w3.org/RDF/>, 2014. Publication date: 2014-02-25; Previous version published: 2004-02-10.
  - [36] W3C OWL Working Group, Web ontology language (owl), <https://www.w3.org/OWL/>, 2012. Publication date: 2012-12-11; Previous version published: 2009-10-27.
  - [37] G. Guizzardi, Ontology, Ontologies and the “I” of FAIR, *Data Intelligence* 2 (2020) 181–191. doi:10.1162/dint\_a\_00040.
  - [38] L. Candela, D. Mangione, G. Pavone, The fair assessment conundrum: Reflections on tools and metrics, *Data Science Journal* 23 (2024) 33. doi:10.5334/dsj-2024-033.