Ontologies for Semantic Annotation: Proposal for an Ontological Multimedia Reference Model

Daniela Lucas da Silva Lemos* and Renato Rocha Souza**

*Federal University of Espírito Santo (UFES). Av. Fernando Ferrari, 514, Vitória, ES, Brazil
**University of Vienna, Universitätsring 1, 1010 Vienna, Austria
*daniela.l.silva@ufes.br / **renato.rocha.souza@univie.ac.at

Daniela Lucas da Silva Lemos holds an PhD in Information Science from the Federal University of Minas Gerais (UFMG). Professor of the Library Science Department of the Federal University of Espírito Santo (UFES). Professor and Researcher at the Department of Library Science and the Postgraduate Program in Information Science at the Federal University of Espírito Santo (UFES), Vitória, Brazil. She coordinates the Laboratory of Information Science, Data and Technology (Cidat Lab) linked to UFES. Holds a Bachelor in Information Systems Administration, a specialization in Strategic Information Management, a Master Degree in Information Science and a PhD in Information Science. She has experience in the area of Information Science, with an emphasis on Information Organization, Knowledge Representation and Information Retrieval, acting mainly on the following topics: Knowledge Organization, Ontologies, Metadata Standards, Semantic Web, Linked Data and Semantic Annotation of Multimedia Resources on the Web. She has experience with IT, with an emphasis on Databases, Software Engineering and Analysis and Design of Information Systems.

Renato Rocha Souza holds an PhD in Information Science from the Federal University of Minas Gerais (UFMG). Professor of the School of Applied Mathematics of the Getúlio Vargas Foundation (FGV / RJ) and Researcher at the Österreichische Akademie der Wissenschaften. Researcher, Professor and Data Scientist with +20 years of experience in Information Science, Applied Mathematics and Computer Science, working with the topics of Data Science, Machine Learning & Analytics; Scientific Programming, Natural Language Processing; Information Management and Retrieval; Knowledge Organization and Representation; Knowledge Management; Education and Research. Holds a Bachelor in Electrical Engineering, two specializations in Technology and Education, a Master Degree in Production Engineering and a PhD in Information Science. Holds two Post Doctorates in Computer Science, and has international experience as a Visiting Fellow (University of South Wales, UK), Visiting Researcher (Columbia University, NY) and Scientific Researcher (ÖAW, Austria).

Lemos, Daniela Lucas da Silva, and Renato Rocha Souza. 2024. "Ontologies for Semantic Annotation: Proposal for an Ontological Multimedia Reference Model". *Knowledge Organization* 51, no. 8: 561-581. 45 references. DOI:10.5771/0943-7444-2024-8-561.

Abstract: Recent years have seen considerable growth of online multimedia databases, largely due to digitization processes in different sectors of society. Knowledge organization and representation strategies were used to qualify and enrich data and metadata from different types of documents and ensure persistent and interoperable online information structures. This study aimed to propose an ontological reference model to systematically organize metadata that describes multimedia documents based on different contexts and needs. The proposed model was based on the NeOn methodology and aimed to encompass the functional and nonfunctional requirements for the construction and reuse of ontology classes obtained by merging and aligning previously analyzed multimedia ontologies. This resulted in a comprehensive conceptualization to organize multimedia metadata for application contexts that deal with the semantic annotation of information entities produced and consumed in the web of data (Semantic Web). We concluded that advances in developing conceptual reference models for representing multimedia documents are the result of interdisciplinary efforts that drive progress in the production and use of more consistent and coherent metadata aimed at facilitating the cross-referencing, interconnection and aggregation of online information sources.

Received: 8 September 2024; Revised: 31 October 2024; Accepted 15 November 2024.

Keywords: Knowledge Representation. Conceptual Modeling. Interoperability. Ontologies. Semantic Annotation. Multimedia Metadata Ontology.

1.0 Introduction

In recent years, digitization has played an important role in enhancing the expansion of large online multimedia databases, as is the case in GLAM (Galleries, Libraries, Archives and Museums) institutions (Europeana Tech 2021; Siqueira and Martins 2021; Lemos et al. 2022; Martins et al. 2022). These institutions use modern mechanisms for scanning heritage assets and contemporary information infrastructures, such as digital repositories, in order to democratize scientific and cultural knowledge on the internet.

Collections of multimedia objects, therefore, have grown considerably in their different types, formats, and complexity, including texts, static, and moving images, videos, sounds, three-dimensional (3D) models, websites, and other specific media. Such resources require different forms of processing and representation to link multimedia documents and improve search, browsing, and retrieval systems using aggregating semantic approaches to web resources.

As such, the growth of digital multimedia objects online is considered impractical in terms of their preservation, location, accessibility, interoperability, and reuse without the support of strategies for information and knowledge organization and representation (Martins et al. 2022, 5) that incorporate good digital curatorship practices, such as maintaining, preserving and adding value to data (Higgins 2011).

Svenonius (2000) reported that information must be described in order to be organized, and information representation is the product of this descriptive process. The author highlights that some types of information representation are constructed through languages, subdivided into languages that describe information (content) and those that describe the document (specific media), either as a whole or in parts.

Language used to describe the document is related to descriptive representation (Gilliland 2016; IFLA 2009; Galeffi et al. 2016; Zeng and Qin 2016), also considered a cataloging process, which involves the creation and use of metadata, making it essential in standardizing and describing information resources that give users the ability to find, identify, select, obtain, navigate, and explore the item inside an online catalog (Galeffi et al. 2016). On the other hand, content description language is associated with thematic representation (Lancaster 1986; NISO 2005), which focuses on intellectual and semantic (subjective) aspects such as understanding the subject of the document for the purpose of translation into a documentary language that helps users select search filters and browse online information systems.

Beyond cataloging principles (Galeffi et al. 2016), of which are geared more towards human users, the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al. 2016; Guizzardi 2020) aim to improve the ability of machines to find digital objects and their metadata on the web via a persistent and unique identifier. They also enable access authentication and authorization, harmonic and effective communication with other applications for different purposes, and comprehensive descriptive information for consumption by both humans and computational agents, with a focus on reuse.

In this context, linked open data (LOD) (Bizer et al. 2009; Machado et al. 2019) stands out as a contemporary technique for organizing and processing documents online and involves using W3C open standards to interlink and annotate data. This allows content providers to enrich their metadata schemas with structured and well-defined knowledge specifications based on standards, vocabularies and ontologies, enabling quality information consumption and reuse.

From this perspective, the W3C recommended metadata standards combined with open data principles (Machado et al. 2019) and quality interoperable data (Guizzardi 2020) have been used to organize and represent multimedia information resources, enabling the expansion of access points and improving the management, organization and recovery of online digital objects. However, there are still few advanced studies on the relationship between multimedia information resources and the Semantic Web on this topic (Ferrada et al. 2018; Lemos and Souza 2020).

Research in the fields of Information Science (IS) and Computer Science (CS) have proposed conceptual models based on semantic technologies for reality-based modeling and the search for and retrieval of information in digital environments (IFLA 2009; Galeffi et al. 2016; Charles et al. 2017; Riva et al. 2017; Fink 2018; Lemos and Souza 2020; Guizzardi 2020; Lemos et al. 2022; Bekiari et al. 2024a; Bekiari et al. 2024b) in order to improve the scope of interoperability between different metadata schemas and applications.

Traditionally, the use of metadata is the most common way of adding semantics to documents (Zeng and Qin 2016); however, the Semantic Web proposes annotating document content using domain ontologies (Shadbolt et al. 2006). In the present study, ontologies are viewed as more sophisticated annotation models (Andrews et al. 2012; Lemos and Souza 2020) in terms of semantic data treatment, allowing users to describe and link existing resources through qualifiers such as the concepts, instances, properties, relationships and constraints between these resources.

Ontologies have been used for the semantic annotation of documents in a variety of applications. For example, in archeology, semantic data models are used to document the geometric aspects of fragments of 3D objects that could be reassembled and reconstructed in specific archaeological research (Catalano et al. 2020); in history, ontological conceptual models are applied to code and disseminate data associated with historic photographic archives (Robledano-Arillo et al. 2020); in the field of digital culture heritage, semantic data models are used for the online publication of cultural collections (Dijkshoorn et al. 2018; Lemos et al. 2022); and in architecture, domain ontologies aim at the reality-based 3D annotation of building conservation state (Messaoudi et al. 2018).

Nevertheless, there are gaps in proposed metadata models for multimedia documents and ontologies for semantic annotation (Van Ossenbruggen et al. 2004; Nack et al. 2005; Lemos and Souza 2020). For example, ISO/IEC standard MPEG-71[1] (Martínez et al. 2002) aims to provide possible solutions for problems associated with producing quality multimedia metadata, but has semantic limitations and the schemas of the different parts that enable descriptions are complex. On the other hand, in terms of proposed ontologies, these generally focus on more generic aspects involving standards, but without considering specific types of metadata for describing multimedia or providing specific descriptors for some types of metadata without relevant modeling reasoning that can ensure, for example, the scalability of their conceptualization. For example, in several of the revised proposals for multimedia ontologies (Lemos and Souza 2020) we expected to find models based on the MPEG-21 Multimedia Framework standard (Kudumakis et al. 2019) for organizing metadata associated with the management of intellectual property rights. This did not occur.

Thus, based on the gaps identified in the above scenario, the present study aims to answer the following opportune and challenging questions: i) *how can a comprehensive conceptual framework that underlies the annotation of multimedia documents be formally expressed?* ii) *what methods and techniques would be suitable for selecting and aligning vocabularies and multimedia ontologies for the annotation of multimedia documents developed by different communities?* and iii) *how can existing types of metadata be systematically organized to annotate multimedia documents for different contexts and needs?*

The present study aimed to advance research on models and modeling for the semantic representation of multimedia documents by proposing an ontological multimedia reference model (OMRM) based on best linked open data practices and on the reuse of existing models in order to cover gaps examined in these, and also expand the coverage of certain important aspects analyzed to systematically organize existing metadata types and describe multimedia documents, according to different contexts and needs. We understand there are advances resulting from the present research for the field of Information Science, especially for the area of Descriptive Cataloging of networked multimedia digital objects, as well as for the area of Ontology Engineering, considering that there are no ontological models for the domain of annotation of multimedia documents that encompass high-level central taxonomic structures, independent of the foundation ontology, and also consider ontological structures based on consolidated ISO metadata standards in library communities digital and multimedia. Additionally, no specific studies were found in the literature that considered categories of descriptive, independent and content-dependent types of metadata represented in reference ontologies of the domain annotation of multimedia documents. Such categories of metadata types are reflected in the way the ontological classes were organized and represented in the OMRM of the present research.

We acknowledge existing ontological models that consider media types and central metadata. However, only for specific domains, such as Europeana's EDM (Charles et al. 2017), IFLA's LRM (Riva et al. 2017; Bekiari et al. 2024b) and ICOM's CIDOC-CRM (Bekiari et al. 2024a), all in the field of Cultural Heritage. The model proposed in this research is intended to be used for multimedia annotations in any domain of knowledge, considering highly relevant multimedia standards such as annotation, decomposition and collection for the realization of the media and the content involved in it.

It is also worth noting that this article is a continuation of research previously published in Lemos and Souza (2020) in which we described in detail the entire methodological path based on the recognized NeOn Methodology, strongly based on reuse on existing models, which led to the careful selection of ontological and non-ontological resources for the OMRM proposed in this article. In this sense, we consider that OMRM is a reference model for the multimedia annotation domain, although it is not yet actually implemented and coded in a computational language, such as Web Ontology Language (OWL) or Resource Description Framework (RDF). However, we consider that the artifact specification delivered in this research already helps in the implementation of OMRM in future research. The OMRM proposed here aims to harmonize (connect) with other vocabularies from any domain available on the network and structured on linked open data principles, including Simple Knowledge Organization System (SKOS), metadata standards with structures in RDF-based languages (e.g.: VRA Core, Dublin Core) and also formal ontologies (e.g.: CIDOC-CRM, DOLCE).

Finally, OMRM's central taxonomic structure, based on a high-level ontology, offers opportunities for harmonization with axiomatized classes that allow structural and semantic alignment with other vocabularies (with or without formal rigor) available for reuse in an open environment.

The above justifications therefore lead us to provide reasons to recommend using the proposed reference model instead of creating an entirely new model.

2.0 Methodology

The proposed OMRM was supported by adopting a current methodological guide, tested and validated for different domains and areas, that follows the guidelines for construct-

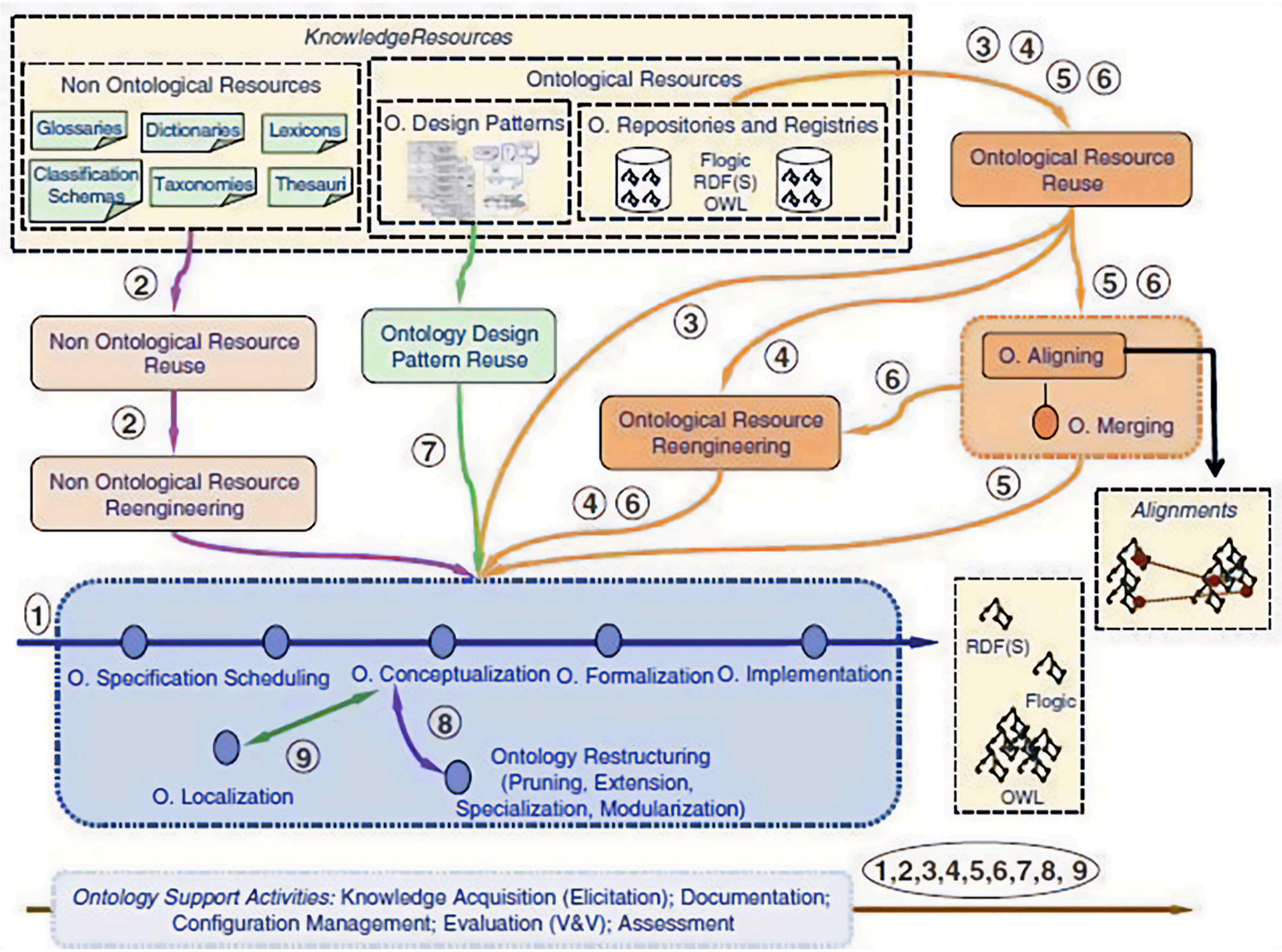

Figure 1. NeOn Methodology Scenario (Suárez-Figueroa et al. 2012, 13).

ing ontology networks based on LOD principles. In this respect, a literature review was carried out in the field of Ontology Engineering and the NeOn Methodology guide was selected from a set of proposals (Silva et al. 2012; Suárez-Figueroa et al. 2012; Falbo 2014; Almeida and Farinelli 2017) because of its LOD practices and the fact that it is the product of methodological frameworks widely accepted in advanced areas such as Software and Knowledge Engineering. Furthermore, some more recent proposals, including the SABiO methodology (Falbo 2014) and the OntoNeo methodology (Almeida and Farinelli 2017), were also derived from the NeOn Methodology.

The NeOn guide covers nine scenarios (Figure 1) that suggest a series of flexible steps for developing ontologies. These scenarios include situations in which ontologies require reengineering, alignment, modularization, localization, support in different languages and cultures, integration with design patterns and non-ontological resources such as metadata standards, dictionaries, thesauri and taxonomies, among others.

Six of the nine scenarios were selected (Scenarios 1, 2, 3, 5, 6 and 8) and are briefly presented in Table 1, which summarizes the results obtained from the methods and techniques applied in the study, including analysis and assessment criteria for ontologies that are candidates for reuse.

The methodological process for these criteria is described in detail in Lemos and Souza (2020, 303-308).

3.0 Results: proposal for an Ontological Multimedia Reference Model

The proposed OMRM was based, a priori, on specifying the requirements for ontologies aimed at multimedia document annotation. The purpose and scope of the model must be defined in order to specify these requirements. The purpose of the model encompasses its intended use, the potential scenarios that require its use and the possible user communities involved in applications that make use of the model, while the scope includes a set of previously determined functional and non-functional requirements.

Stages of the modeling process	Use of NeOn scenarios	Results generated
Identify and select ontological and non-ontological resources in the **multimedia document annotation** domain.	**Scenario 2:** reusing and reengineering non-ontological resources. **Scenario 3:** reusing ontological resources.	**Scenario 2**: **parameter elements** (120 in total), that is, a set of multimedia features based on the MPEG-7 and Dublin Core metadata standards. **Scenario 3**: **ontologies selected** (9 in total) via the literature review and Semantic Web repositories.
Analyze and compare multimedia ontologies according to previously proposed requirements.	**Scenario 3**: reusing ontological resources.	**Weighted ranking** of candidate ontologies for reuse.
Select appropriate multimedia ontologies to reuse knowledge resources for the construction of the proposed model.	**Scenarios 3 and 5**: reusing, aligning and merging ontological resources	**Arrangements** to **organize knowledge resources** according to the types of metadata addressed in the study.
Develop an ontology-based conceptual model for the **multimedia document annotation** domain.	**Scenario 1:** specifying functional and non-functional requirements; and proposing the conceptualization of the ontology. **Scenario 6**: reengineering aligned and merged ontological resources. **Scenario 8:** restructuring the conceptual model to meet the established requirements.	**Class diagrams of the Ontological Multimedia Reference Model**

Table 1. Summary of the results generated by the methodology applied in the study.

The purpose of the OMRM is to represent a consensual conceptualization shared by a given community for the semantic organization of annotations aimed at multimedia documents that are produced, described, published and consumed online, along with their annotations or metadata. Thus, the conceptual model seeks to enrich different types of multimedia metadata via an information framework suited to scenarios that involve, for example, data aggregation in LOD environments, enabling syntactic and semantic interoperability between different institutions and their information systems.

The OMRM can be used, for example, as a key element in the information systems of cultural heritage institutions, whose users consume, interpret, manipulate and generate multimedia content in their collections that are generally digitized and accessible in online digital repositories. The content of these collections archives can be mapped for the OMRM, whose semantic structure incorporates the notion of an Event (a key entity in the cultural domain), making it possible to aggregate actors, objects (physical and abstract), locations and the duration of time intervals. For example, a historical image could be modeled as a sequence of chronological lines containing persistent items (objects and people) combined in events within a time period. This improves the semantic enrichment of data in terms of information retrieval, since data on related (semantically aggregated) events can be collected to create a powerful semantic network of biographical and contextual data on people, documents, objects and places, which would be useful for educational and scientific research. In short, the OMRM makes it possible to classify elements referenced in cultural heritage documents into formal categories, producing legible descriptions of events and objects that improve search, navigation and retrieval systems by aggregating semantic approaches to data from heritage collections.

The OMRM can also be used to expand possibilities for research and collaboration between users in semantically linked knowledge networks on the Semantic Web, including sources of medical, cultural, multimedia, artistic, historical, tourism, educational and social media-related information such as Wikidata, Wikimedia Commons and Wikipedia (Mora-Cantallops et al. 2019; Navarrete and Villaespesa 2021). This would enhance the exploitation of information in integrated knowledge networks, as well as the circulation and collaborative production of information resources that are useful to society. Catalogers are another category of users that play an important role in linking annotations in multimedia documents, particularly in dynamic knowledge spaces such as digital repositories and libraries. Finally, but by no means the last possible use for the model, a variety of news websites require efficient methods for organizing multimedia content and transmitting it intelligently to different types of users.

3.1 OMRM requirements

The scope of the OMRM was determined based on functional (FR) and non-functional requirements (NFR) established according to the results of comparative analyses of candidate ontologies for reuse (Lemos and Souza 2020). It is therefore relevant to elucidate FR and NFR within the scope of this study. The former encompasses Software Engineering practices that have been adapted to the field of Ontology Engineering (Fernández-López et al. 1997) in order to facilitate tasks involved in specifying the content of a particular knowledge domain, obtaining ontology-related terminologies. In both fields, NFR are use-related requirements that include performance, usability, reliability, security, availability, maintainability and technologies involved.

It is noteworthy that for FR (1 to 4), competency questions were described that reflect important features of the domain investigated here and any application context that deal with multimedia objects. The competency questions method (Grüninger and Fox 1995) involves determining and applying a series of types and examples of questions formulated in natural or formal (first-order logic) language and empirically designed to be efficiently and correctly answered by the ontology. These questions reflect the main knowledge demands of future users of the ontology. This method enables a practical and intuitive description of the ontology's requirements and scope, helping to obtain a more accurate perspective of the classes, properties and relationships that must be included (Robledano-Arillo et al. 2020). These requirements are described below.

- (FR1) Covers content independent metadata: these data are not directly related to media content, but used to manage and administer information resources, such as creation and production, genre, language, format, usage rights and age-restricted content, among others. Examples of competency questions: i) *where are images created?;* ii) *what age range can access a given program?;* iii) *what are the resolutions of the image files?;* iv) *what are the copyrights of a user related to a certain media (including its content)?*; and v) *which agent is responsible for publishing a copy or part of a work?*
- (FR2) Covers content dependent metadata: the features of visual and audio data are considered primitive or low level in that their content, such as color, texture, shape, spatial relations, movement, location, spectral and temporal timbre and signal parameters, are generally extracted automatically by computer algorithms. Examples of competency questions: i) *what is the predominant color of an image?;* ii) *what part of an audio stream is predominated by the timbre of a musical instrument?; and* iii) *what are the geographic coordinates of an object located in a city shown in an image?*
- (FR3) Covers descriptive metadata: these refer to semantic content that links media entities with their real-world counterparts, such as the face of a person portrayed in an image, as well as aspects involved in personalizing content to facilitate navigation, access and user interaction in relation to content consumption. Examples of competency questions: i) *which frames of a video depict a certain scene?;* ii) *which chapters of a book address a specific subject?;* iii) *which spoken documents portray a particular statement?*
- (FR4) Considers media content and realization in different formats, such as audio, image, text, 3D models and video: separating information objects from their realizations is important in terms of easily visualizing different manifestations (a book in PDF format, a 3D digital replica, songs recorded in an MPEG file) of the objects (a story, a sculpture, a song) and clearly understanding the relationships between them and their realizations. Content-independent metadata such as file size or media location on the web are typically applied to information realization, whereas descriptive metadata for multimedia content aims at describing the message to be conveyed to the consumer of the content. As such, this separation is relevant in that it provides a clear distinction between content semantics and the data itself (e.g. media file). Examples of competency questions: i) *what versions are there for a specific multimedia presentation available online?;* ii) *what are the examples of a literary work in its multiple expressions and manifestations in a bibliographic collection?; and* iii) *what photographs and manuscripts are available in the archival collection for a specific theme, agent or object?*
- **(NFR1) Has an upper ontology as reference:** upper ontologies are referred to as *foundational ontologies* (Guizzardi 2020) and describe very broad concepts such as space, time, matter, objects, events, agents etc. They are considered philosophically well-founded systems of domain-independent categories. Their use semantically benefits the core taxonomy of the domain ontology by clarifying the intended meaning of the terms, supporting, for example, the integration of instances of media content with domain ontologies.
- **(NFR2) Is based on extended multimedia patterns with an LOD approach:** mitigates the challenges of reuse with acceptable and memorable diagrammatic visu-

alizations for a specific set of competency questions (problem and its solution). Some ontologies for multimedia annotation use design patterns to generically organize entities and relationships underlying the multimedia domain, such as annotation and decomposition. These patterns make it possible to link different media resources and coherently integrate metadata (annotations) involved in these resources via a URI, in line with LOD principles. Thus, NFR2 works in conjunction with NFR1 in that the patterns inherit associated axioms and inference services from their upper ontology.

– (**NFR3**) **Considers well-placed ontologies in a ranking** produced from well-founded criteria for reuse: use of a mature, robust and efficient methodology for careful analysis and evaluation of ontologies for multimedia annotation (Lemos and Souza 2020). Reusing available knowledge resources to model knowledge of a domain is recommended in the field of Ontology Engineering (Fernández-López et al. 1997).
– (**NFR4**) **Considers different levels of granularity**: provides a conceptual model that represents a comprehensive taxonomic structure capable of supporting generic (e.g. annotation) and specific multimedia entities (e.g. primitive and specific audio descriptors) according to a particular context.
– **(NFR5) Ensures interoperability** in relation to multimedia content on the web: ensures that the intended meaning of the captured semantics can be shared between different applications within the scope of the Semantic Web. In addition to exchanging multimedia content, the model should also provide the means of transmission in a syntax agreed upon by a community which, in this case, would be via Semantic Web languages such as RDF/OWL.
– **(NFR6) Has an architecture that allows separation of concerns:** provides clear separation of concerns involving the subject of the media (content semantics), knowledge related to managing media information resources (content-independent metadata), structure (media segments) and features of multimedia documents (content-dependent metadata).
– (**NFR7**) **Has an extensible architecture** in terms of building a comprehensive multimedia ontology: since an ontology is always evolving, the inclusion of new concepts must be stipulated in the extensible conceptualization. Extensibility is ensured to the extent that design patterns and upper ontologies can, through meta-categories, expand the possible insertion of new concepts without needing to change the core underlying model.

3.2 Selection, features and alignments of multimedia ontologies suitable for reuse

The comparative analysis of the ontologies for multimedia annotation suitable for reuse was carried out using criteria organized into 4 dimensions (Lemos and Souza 2020, 305-312), as follows:

i) Resource Reuse Effort: estimation of costs related to time and economy required to reuse the evaluated ontology;
ii) Resource Understandability Effort: estimation of effort required to understand the content of the evaluated ontology;
iii) Resource Integration Effort: estimation of efforts undertaken to integrate the evaluated ontology to the new ontology that is being built; and,
iv) Resource Reliability: analysis of the performance of the ontology evaluated against aspects of semantic treatment in declarations (e.g., axioms present, knowledge resources used), evaluation (e.g., available tests) and renowned projects that make use of them.

Weighted ranking (Figure 2) and the findings of comparative analysis of candidate ontologies for reuse enabled the selection and justification of ontological resources according to the previously described requirements. The ontologies best suited to the proposed OMRM were the Media Ontology (1.56), M3O (1.23), COMM (1.19) and M3 Multimedia (0.95). Most of these ontologies use the MPEG-7 metadata standard to support their constituent elements.

The Media Ontology was proposed in 2009 by members of the W3C Media Annotation Working Group, which aims to improve interoperability between metadata schemas for web-based media resources, such as video, audio and images.

The M3O or Multimedia Metadata Ontology (M3O) was created in 2010 as a comprehensive model to represent metadata aimed at multimedia document annotation, including combinations of ontological models commonly used in the Semantic Web.

The COMM or Core Ontology for Multimedia was developed in 2007 by a group of renowned researchers in the fields of multimedia, digital libraries and the Semantic Web. Its main purpose is to provide a sound conceptualization, based on the MPEG-7 metadata standard, that broadly covers a specific domain dealing with multimedia content.

M3 Multimedia was created in 2012 as part of a comprehensive ontology (addressing different domains and languages) denominated the M3 Ontology Network, the product of a Spanish research project involving entities such as the Ontology Engineering Group.

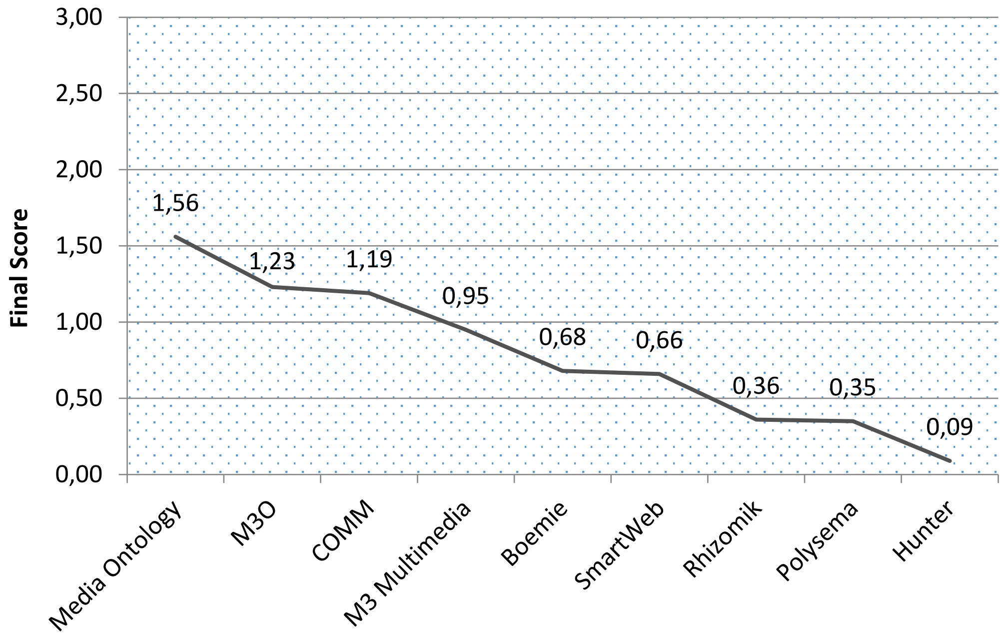

Figure 2. Weighted ranking of candidate ontologies for reuse (Lemos and Souza 2020, 312).

Given that the OMRM must be based on an upper ontology (NFR1) and multimedia design patterns (NFR2), in addition to addressing semantic differences between media content and realization (FR4), M3O was selected as the ontology that best met these requirements and was therefore considered central in the model. This choice is justified by the fact that M3O's conceptualization architecture is based on the upper DOLCE+DnS Ultralight (DUL) ontology (Masolo et al. 2003; Borgo and Masolo 2009) and three design patterns that it references, as described below: Descriptions and Situation (DnS), Information and Realization Pattern and Data Value Pattern.

The *Descriptions and Situation* pattern provides an ontological formalization of context based on role assignment. The semantics embedded in this design pattern (see the application example in the class diagram in Figure 6) state that a situation satisfies (the satisfies relation) a description in which a situation is a specific context with concrete entities that express a certain role; and the description, in turn, is a conceptualization that defines certain concepts which determine (the classifies relation) the roles of entities in a specific context. Entities, on the other hand, are considered relevant or true only in a given context.

Each entity is connected to a given situation via the hasSetting relation, such as quantitative metadata (represented by an entity) related to a color histogram participating (hasSetting relation) in image annotation.

The Information and Realization pattern (Gangemi and Presutti 2009) models the distinction between information objects and information realizations, underpinning FR4. In the example in Figure 3, the class related to the information realization formally realizes some information object (with its inverse is realized by relation). Both concepts are InformationEntity subclasses that allow information to be treated in a general sense (see the application example in the class diagram in Figure 4).

The *Data Value* models the concrete values of an entity in order to reduce the risk of ambiguities (Saathoff and Scherp 2010). DUL (see application example in the class diagram in Figure 4) contains the concepts of *Quality and Region*, which represent, respectively, the intrinsic attributes of an entity linked to its values with corresponding data spaces. In pattern description, the attribute is represented by the concept *Quality*, which is connected to the Entity by the *hasQuality* relation. *The Quality* is connected to a Region via the *hasRegion* relation, and the concrete value is attached to the *Region* by the *hasRegionDataValue* relation. The primitives used in the *Data Value* pattern are useful in expressing structured data values supported by MPEG-7, especially for data automatically extracted from media, such as color, texture and shape, among others.

M3O multimedia patterns are therefore extended from the *Descriptions and Situation* Pattern, including *Annotation Pattern*, *Decomposition Pattern* and *Collection Pattern*, as presented in the next section. Their design diagrams are easily recognizable by the simplicity of their few class and relationship schemes, making it possible to understand the modeling reasoning used in conceptualizations. Addition-

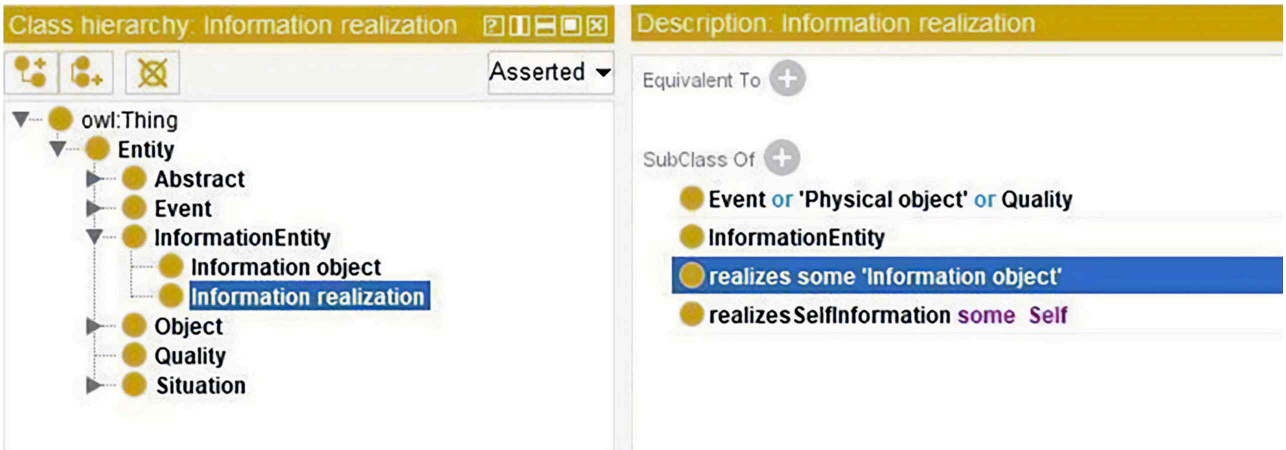

Figure 3. Description of DUL information objects and information realization
Source: screenshot of Protégé software used in the content analysis of the ontology DUL.

ally, all three multimedia patterns (annotation, decomposition and collection) act on the semantics specified in the *Information and Realization* pattern, enabling objects of information and their realizations to be annotated, decomposed and organized into collections.

Media Ontology is recommended for content-independent metadata (FR1) because of its satisfactory coverage index (Lemos and Souza 2020, 313) for this type of metadata (in relation to the other ontologies analyzed), particularly for descriptors aligned with the Dublin Core metadata standard.

The COMM and M3 Multimedia ontologies provide descriptors that align well with content-dependent metadata (FR2), useful in computer processing of digital data for the automatic generation of quantitative metadata. Both have very similar visual coverage indices (Lemos and Souza 2020, 310-311), particularly for descriptors involving color, texture, shape and location of regions of interest. Metadata to describe 3D characteristics, for example, are present in both ontologies, primarily for shape-related visual aspects, since both are based on MPEG-7 for multimedia content description. MPEG-7 includes descriptors for the geometric characteristics of 3D objects, such as symmetry, circularity, axis location, size and orientation of consecutive border segments, curvature points and angles of curves. Knowledge resources related to audio metadata can be selected from M3 Multimedia because it reuses both the visual and audio metadata from the VDO Boemie ontology (Lemos and Souza 2020).

Descriptive metadata (FR3) aimed at the semantics of media content are generally linked to instances of domain ontologies or controlled vocabularies with less formal rigor (called Simple Knowledge Organization System - SKOS) whose semantic labeling is organized within the taxonomy of an upper ontology. Since M3O is part of DUL, it plays the role of organizing semantic labels from domain ontologies or SKOS into abstracts entities such as event, object, time, place, etc., in addition to dealing with their relationships. M3 Multimedia covers properties for navigation and access (content customization), audio descriptors with high-level features (spoken content, for example), and common descriptors for segment annotation.

It should also be noted that because M3O uses an upper ontology as reference and is based on multimedia patterns extended from ontology design patterns, all its features meet the previously outlined nonfunctional requirements, such as interlinking open license data (media and its metadata), treating different levels of granularity, interoperability, separation of interests and extensibility. The possibility of linking different media resources and integrating metadata is achieved by a semantic URI that uniquely identifies the resources (entity at its most abstract level) in the network. The data value corresponding to the URI can be modeled through the Entity in DUL (rdfs: domain primitive) by setting any URI (rdfs: range primitive).

3.3 Arrangements and mappings of the OMRM ontology classes

Following the alignment of ontologies with the predetermined requirements, as elucidated in the previous section, knowledge resources were semantically organized into proposed arrangements or groups, including: i) information objects (document content) and their realizations (media) involved in the annotation context; and ii) types of multimedia metadata (particularly from MPEG-7) and their respective ontology classes based on content-dependent, content-independent and descriptive metadata. That said, the

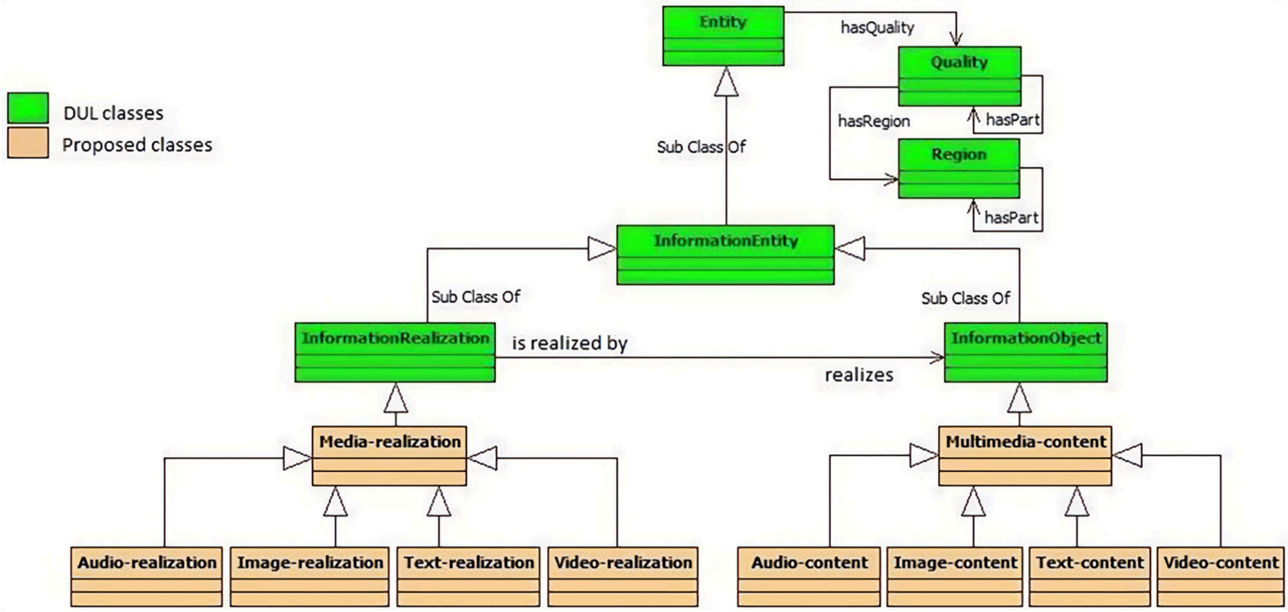

Figure 4. Classes of key entities in the Ontological Multimedia Reference Model.

mappings were determined from the arrangements and multimedia patterns proposed for the model.

In the first arrangement (Figure 4), information objects and their realizations were grouped into the *Multimedia-content* and *Media-realization* ontology classes, respectively. These classes were then generalized to the DUL classes *Information object* and *Information realization*, respectively.

New media specializations can be inserted into the OMRM taxonomy for specific applications, given the expansible nature of the model.

Tables 2 and 3 present the remaining arrangements organized into hierarchical structures for organizing knowledge resources according to the types of multimedia metadata addressed in the study. In spite of these tables presenting informal definitions, they are considered important tools to support communication, in addition to enabling the discussion, negotiation and representation of the consensus reached among domain specialists for future implementations involving specific domains that deal with multimedia metadata.

We opted to map the taxonomic structure of the ontology by identifying its subclasses and superclasses. The "." symbol (point) indicates the subclass relationship between the classes involved; for example, color as a subclass of visual metadata. Corresponding ontology classes were mapped (or proposed) for each ontology involved in reuse and for some concepts from MPEG-7.

It is important to underscore that the nomenclature of the classes was maintained according to their ontological origin, whereas the proposed classes were named in line with naming conventions underlying the use cases studied. The (*) symbol signals a new class for the model.

Ontology models based on design patterns (such as COMM and M3O) already contain formal coherent groupings (provided by the axioms of their upper ontologies) represented in specific ontology classes. For example, the COMM *localization-descriptor-parameter* class contains the concepts *region-locator-descriptor-parameter, bounding-box* and *region-boundary*, which correspond to metadata for visual localizations.

On the other hand, the properties of ontology models focused on relations and attributes (such as Media Ontology and M3 Multimedia) are grouped into ontology classes. In these cases, axiom modeling is necessary in order to formally establish the metadata elements of the groupings as a specific type. Constraints can be modeled using logical statements including existential quantifiers to indicate relationship to at least one individual, and universal to indicate relationship to all individuals. A constraint could be modeled, for example, for the Media_Creation class (shown in Table 2) declaring that there is a unique identification (class identification) and a location where the resource can be accessed (class locator) associated with an instance of the entity class with media creation and production roles. Figure 5 displays a code example representing such an axiom.

Other types of axioms can be created, for example, to restrict the participation of metadata in the annotation of an information entity, considering the semantic distinction between an information object and an information realization, subjects of the annotation. In the example in Figure 5, the locator metadata from the Media Ontology has the role

Type of metadata	Ontology Class	Class Description
Content-Independent Metadata		
Media Ontology		
Type of metadata	**Ontology Class**	**Class Description**
Media creation and production	Media_Creation (*)	Describes the features involved in creating media content and its associated resources.
Media classification	Media_Classification (*)	Describes features aimed at classifying media, such as genre, subject, purpose and language, among others.
Media information	Media_Information (*)	Describes types of storage, including content format, compression and coding.
Media usage	Media_Usage (*)	Describes features that reflect the usage rights, registration and availability of media usage.
Content-Dependent Metadata		
COMM Ontology		
Type of metadata	**Ontology Class**	**Class Description**
Visual	structured-data-parameter.visual-descriptor-parameter	Describes primitive visual features for color, texture, shape and motion.
Color	.color- descriptor-parameter	Describes various descriptors and supporting parameters in the representation of different aspects of color-related features.
Texture	.texture-descriptor-parameter	Describes important aspects in revealing tactile, depth and surface orientation features for an image.
Shape	.shape-descriptor-parameter	Describes features related to the spatial arrangement of pixels that belong to an object or region. The descriptors can be grouped into 2D or 3D classes.
Motion	.motion-descriptor-parameter	Describes spatial and temporal features captured by camera movement, a moving object, or both.
Localization	localization-descriptor-parameter	Describes localization for regions of interest in spatial and spatiotemporal domains.
M3 Multimedia Ontology		
Type of metadata	**Ontology Class**	**Class Description**
Audio	LL_Audio_Descriptor	Describes primitive descriptors involving spectral, parametric and temporal features to describe audio signals and files.
Spectral basis	.Spectral_Basis_Descriptor	Describes low-dimensional projections of a high-dimension spectral space to aid in compactness and identification.
Spectral timbre	.Spectral_Timbral_Descriptor	Describes timbre features related to the signal spectrum.
Temporal timbre	.Temporal_Timbral_ Descriptor	Describes temporal features of audio segments; particularly useful in describing the timbre features of musical instruments.
Signal parameters	.Signal_Parameter_ Descriptor	Describes periodic or quasi-periodic signals.
Basic spectral	.Basic_Spectral_Descriptor	Describes descriptors derived from signal frequency analysis.
Basic	.Basic_Descriptor	Describes basic descriptors for general use and applicable to all types of signals.

Table 2. Arrangements for types of content-dependent and content-independent metadata.

of annotating media files that are located on the Web. Such an assignment should only apply to one realization of the information and, therefore, should be formally declared in the locator annotation class. On the other hand, the description metadata (also from the Media Ontology) has the role of annotating media content information. Thus, this metadata should only apply to information objects, and therefore an axiom should be created for the description annotation class to enforce such a restriction.

Following the proposed arrangements involving the knowledge resources of the study (indicated in Tables 2 and 3), the ontology classes of the M3O multimedia design pat-

Descriptive Metadata			
M3 Multimedia Ontology			
	Type of metadata	**Ontology Class**	**Class Description**
	Navigation and access	Navigation_Access (*)	Describes aspects of features that facilitate navigation and access to multimedia content, such as summaries.
	High-level audio	HL_Audio_Descriptor	Canonically describes a sound with a certain degree of generality, including descriptors aimed at covering specific domains.
	Spoken content	.Spoken_Content_Descriptor	Describes details of spoken words in an audio stream.
M3O Ontology			
	Type of metadata	**Ontology Class**	**Class Description**
	Organization of digital objects into collections	CollectionPattern	Describes features of collections of information entities with common properties.
	Media segments	DecompositionPattern	Describes the structure of multimedia content in terms of segments, such as frames, moving and static regions and audio tracks.
	Content semantics	DUL:Entity	Describes real-world objects, events and notions that can be abstracted from multimedia content.
MPEG-7 metadata standard			
	Type of metadata	**Ontology Class**	**Class Description**
	Temporal segment	Temporal_Segment (*)	Describes a set of temporal features related to segment decomposition for specific media content, such as video, audio, scenes and moving regions.
	Spatial segment	Spatial_Segment (*)	Describes a set of spatial features related to segment decomposition for specific media content, such as 2D and 3D images and moving regions.
	Spatiotemporal segment	Spatio_Temporal_Segment (*)	Describes a set of spatiotemporal features related to segment decomposition for specific media content, such as moving and audiovisual regions.

Table 3. Arrangements for types of descriptive metadata.

terns (*Annotation Pattern*, *Decomposition Pattern* and *Collection Pattern*) were semantically mapped, as shown in the class diagrams described below.

In order to ensure better visualization and understanding, the OMRM was segmented into three parts associated with the multimedia design patterns underlying the proposed conceptualization. For the purpose of easy visualization, the diagrams depict more general as opposed to specific classes.

In M3O, an *AnnotatedConcept* classifies an *InformationEntity* that is the information resource (physical or digital object) to be annotated (as a whole or in parts). Each metadata element is represented by an *Entity* (with a semantic URI) classified by an *AnnotationConcept*. The mappings resulting from this conceptual framework model an information entity (which may be an information object or information realization) and the metadata that participate in the annotation process.

The classes (derived from groupings or mapped from the corresponding ontologies) referring to multimedia metadata were specialized (*Subclass Of* relation) in the *AnnotationConcept* class (Figure 6), which assigns the data entities the role of annotation and formally describes their nature as metadata. For example, an image object (information object) that requires annotation by a semantic concept from a domain ontology (eg.Wikidata) would be classified by an *AnnotatedConcept*. The domain ontology instance that plays the role of semantic metadata would be classified as an *AnnotationConcept*. The link between the image object and the instance of a Wikidata semantic structure is established by the *hasSetting* relation, whereby, as a rule, all the DUL *Entities* (event, object, agent, place and time) has a '*hasSetting'* with the annotation situation that satisfies the *AnnotationPattern*.

The class diagram in Figure 6 shows semantic mapping for the previously described annotation pattern.

Figure 7 depicts the taxonomic structure of the annotation pattern, indicating the axiomatization of the AnnotationSituation class.

Mappings involving the Decomposition Pattern are ensured by the media type classes aligned with the DUL information entities, as shown in Figure 4. As such, a *Composite-*

```
<owl:Class rdf:ID="Media_Creation">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:someValuesFrom>
            <owl:Class rdf:ID="identifier"/>
               </owl:someValuesFrom>
                 <owl:onProperty>
           <owl:ObjectProperty rdf:resource="&DUL;'has part'"/>
       </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
        <owl:Restriction>
          <owl:onProperty>
            <owl:ObjectProperty rdf:resource="&DUL;'has part'"/>
               <owl:someValuesFrom>
                 <owl:Class rdf:ID="locator"/>
               </owl:someValuesFrom>
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
```

Figure 5. Example of restriction for the Media_Creation class of the proposed model.

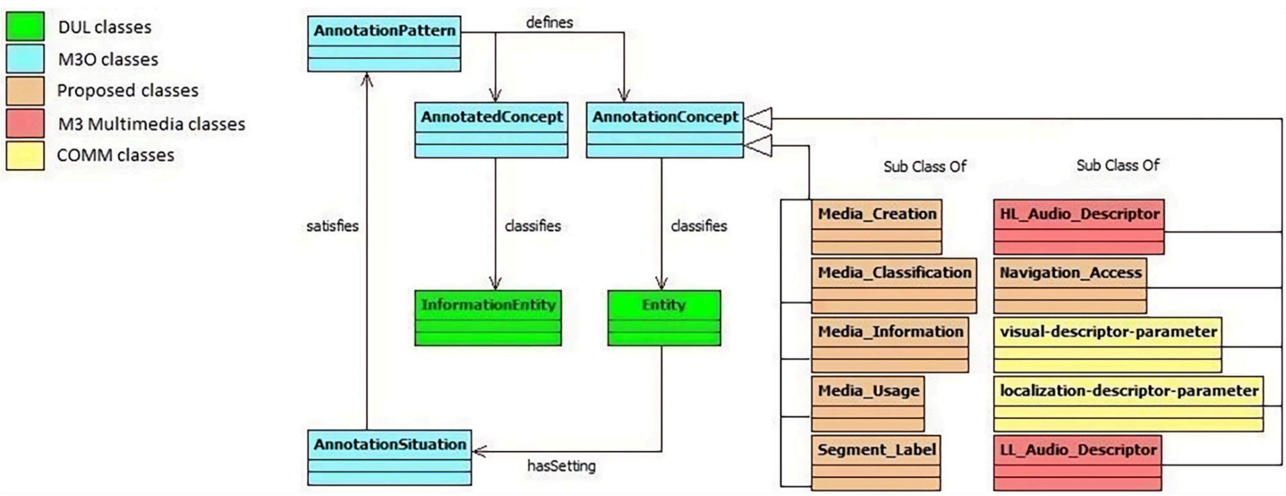

Figure 6. Classes of annotation entities in the Ontological Multimedia Reference Model.

Concept plays the role of an object or document involved in decomposition and the *ComponentConcept* that of the segments resulting from its decomposition. Both classes are configured with a semantic URI inherited from the DUL *Entity class*.

In the case of *CompositeConcept*, the use of the *owl:disjointWith* constructor to limit the participation of instances of media types in inappropriate classes is crucial. For example, an *Audio-content* is disjoint from *Video-content* and *Image-content*. For *ComponentConcept*, the classes proposed for segment types resulting from decomposition were specialized (*Sub Class Of* relation) as *ComponentConcept* subclasses named *TemporalSegment*, *SpatialSegment* and *SpatioTemporalSegment*, representing temporal, spatial and spatiotemporal features of dimensions, respectively.

For these classes, axiom modeling is important in order to impose restrictions on the segment types that form valid decompositions for content involving specific media. For example, a segment corresponding to a moving region (specialized from the *SpatioTemporalSegment* class) would be classified only as video media, and its localizations aimed only at annotation classes involving metadata on visual and time localizations.

This highlights the participation of the annotation pattern in describing the resulting segments for the type of metadata involved, including information on the access location, creator, media, and usage license of the segment.

The class diagram in Figure 8 shows semantic mapping for the previously described decomposition pattern.

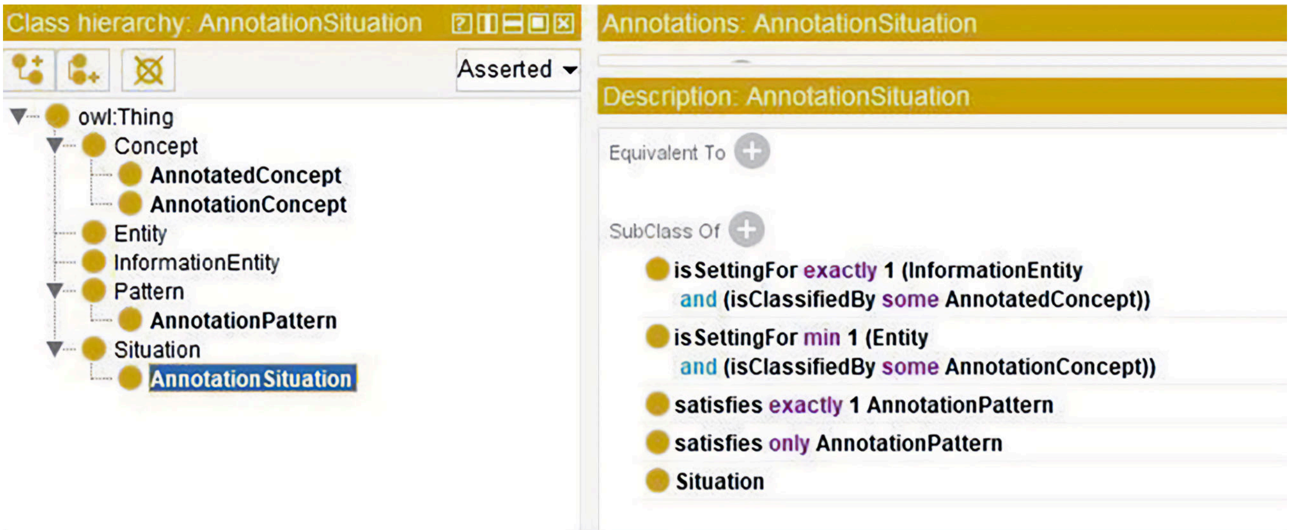

Figure 7. Annotation of information entities.
Source: screenshot of Protégé software used in the content analysis of the ontology M3O.

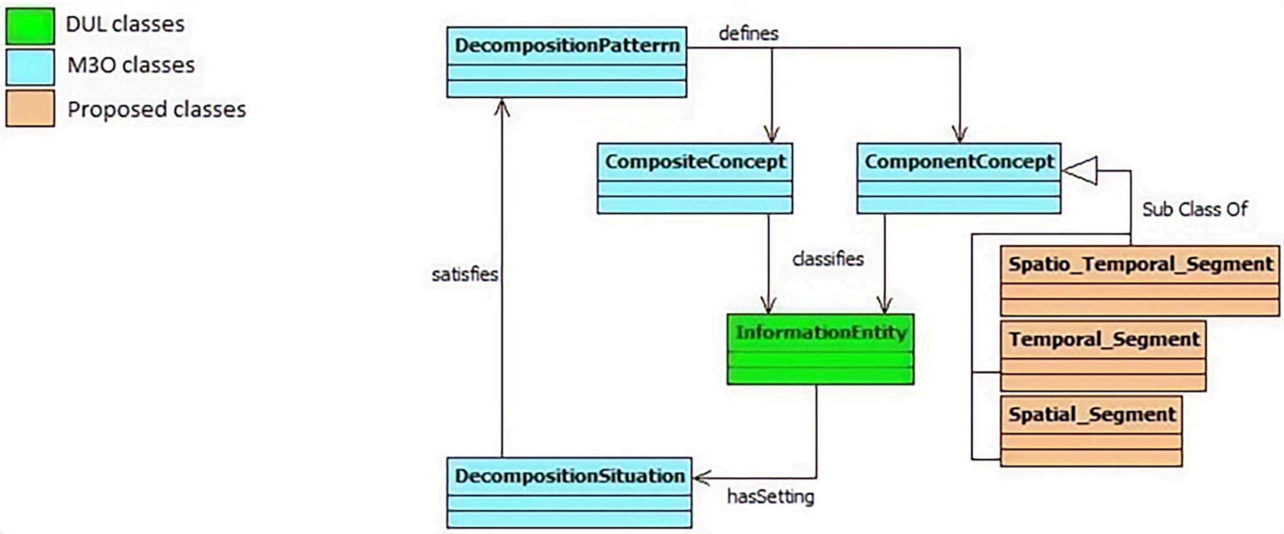

Figure 8. Classes of decomposition entities in the Ontological Multimedia Reference Model.

Figure 9 shows the taxonomic structure of the decomposition pattern, indicating the axiomatization of the *DecompositionSituation* class.

The multimedia pattern for the M3O collection makes it possible to represent collections of information entities with common properties via *CollectionPattern*, which supports the collaborative creation of collections by taking the source or origin of the information entities involved into account. A set of images collected by different people about a common subject is an example of a collection, which can be shared on an online community such as *Flickr*.

The core concepts of the collection pattern establish specializations with DUL design patterns. The *CollectionPattern* class stipulates the existence of exactly one *CollectionConcept* that classifies an *InformationEntityCollection*, which, in turn, is a collection of information entities. Finally, *AnnotationPattern* is integrated and interacts with collection pattern classes in that it provides classes of multimedia metadata to annotate entities in *InformationEntityCollection*.

The class diagram in Figure 10 shows semantic mapping for the previously described collection pattern.

Figure 11 depicts the taxonomic structure of the collection pattern, indicating the axiomatization of the *CollectionSituation* class.

After semantic mapping, the following tasks are recommended, based on the ontology engineering methodologies

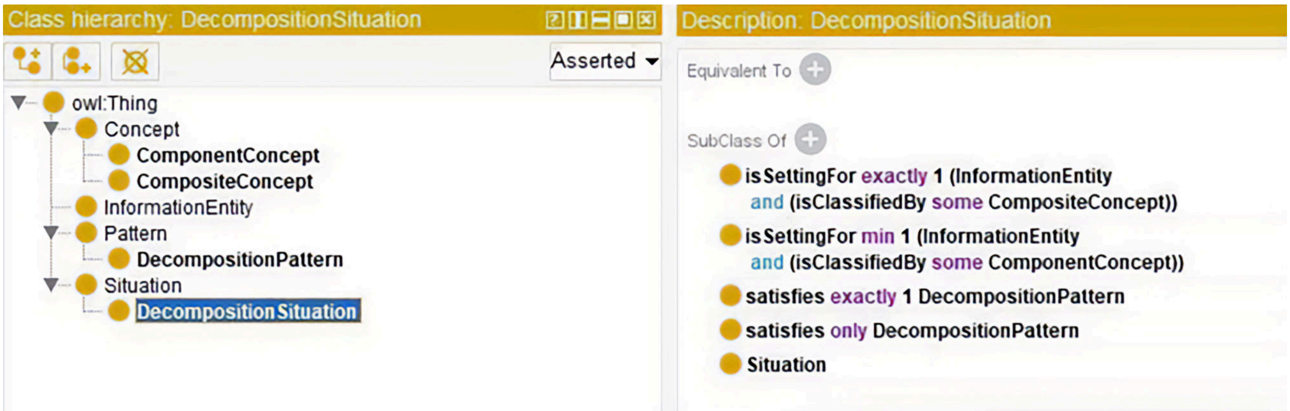

Figure 9. Decomposition of information entities.
Source: screenshot of Protégé software used in the content analysis of the ontology M3O.

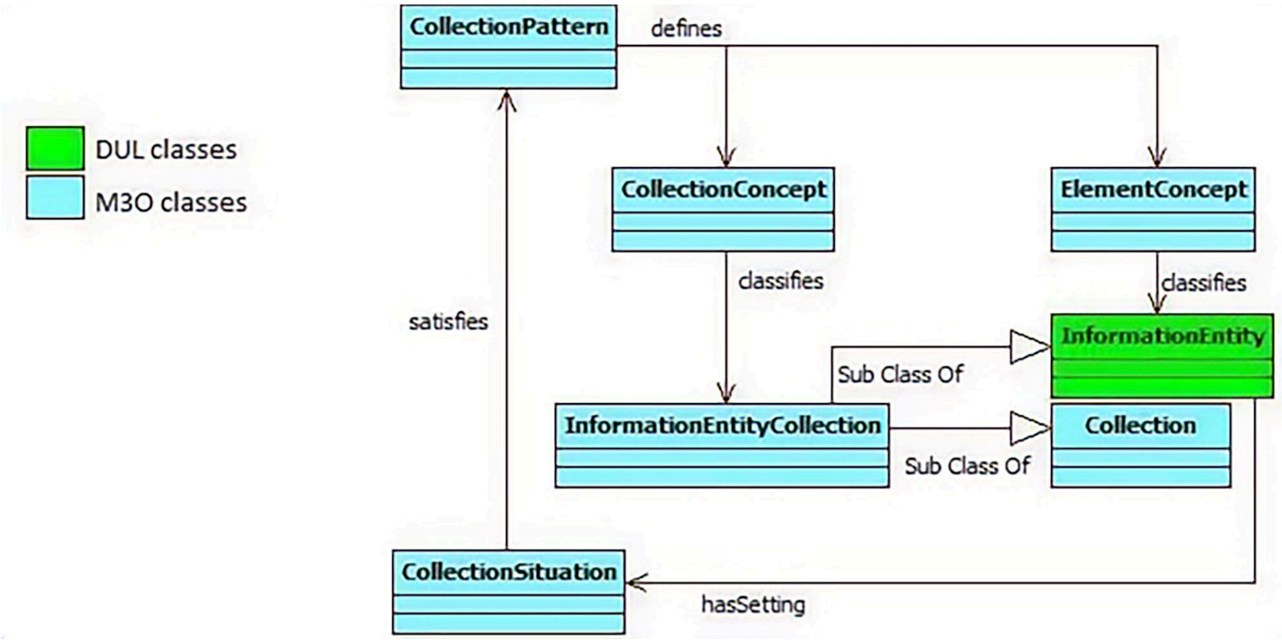

Figure 10. Classes of collection entities in the Ontological Multimedia Reference Model.

operationalized during the study: i) remove unnecessary concepts from the conceptualization of the resulting ontology to prevent an extensive taxonomy with ambiguous concepts; ii) clearly and accurately document all the ontological elements in the conceptual reference model; and iii) validate the taxonomy to assess the consistency of the resulting ontology.

Finally, experts who deal with multimedia files should be consulted to ensure that the domain can be modeled satisfactorily. For example, Library and Information Science professionals are experts in descriptive and subject cataloging of information resources and can contribute by establishing agreements on specific metadata for each information entity involved in model specification. Computer vision, image processing and audio signal experts can contribute to modeling decisions regarding content-dependent metadata, which are heavily dependent on technical knowledge in these areas. This allows ontology engineers to focus on tasks related to modeling domain knowledge.

4.0 Discussion

The proposed ontological model can be characterized as a reference in conceptual specification for multimedia documents that specifically target internal curatorship in information systems aimed at designing normalized and enriched open databases, with a view to improving contemporary information retrieval systems (Mora-Cantallops et al. 2019;

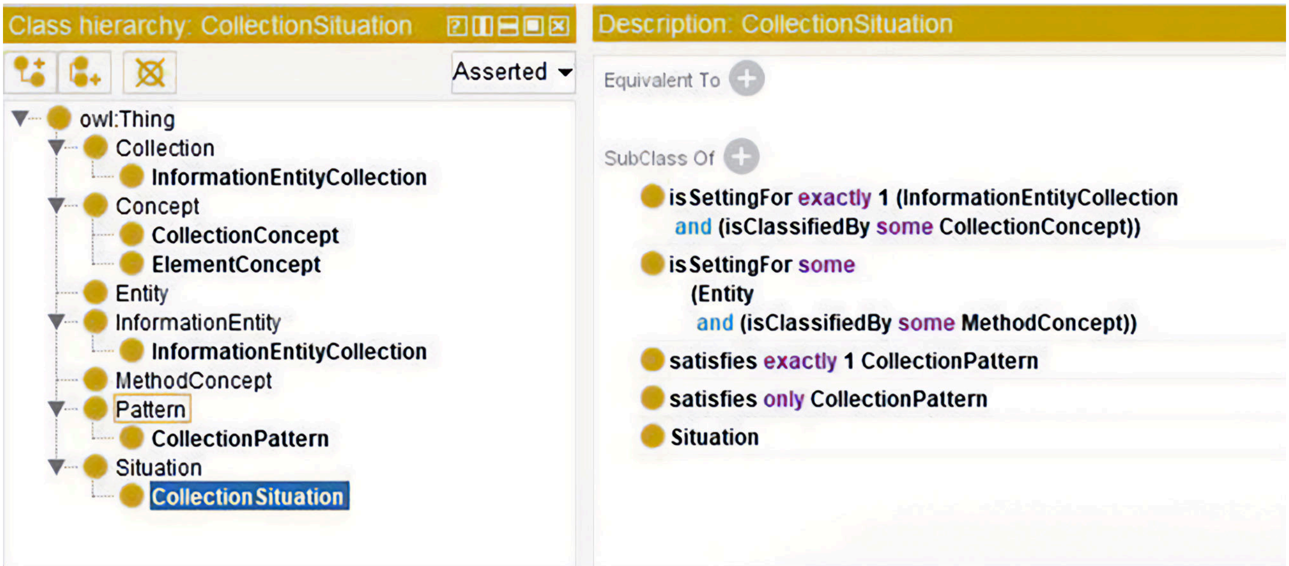

Figure 11. Core concepts of the collection pattern.
Source: screenshot of Protégé software used in the content analysis of the ontology M3O).

Navarrete and Villaespesa 2021; Siqueira and Martins 2021) by providing useful metadata that enable the discovery, reuse, aggregation and integrated search for multimedia objects online.

Metadata standards can be considered the product of the historical development of bibliographic standards and are therefore linked to cataloging codes, conceptual models and new methodological elements for managing information, such as Resource Description and Access (RDA) and FAIR and LOD principles in different online media (Zeng and Qin 2016; Martins et al. 2022). As such, they play a vital role in describing information resources, resulting in qualified access points for discovery, search and retrieval.

However, over the course of the study, significant problems were identified in relation to multimedia metadata (Van Ossenbruggen et al. 2004; Lemos and Souza 2020), mainly due to the convergence of information processes on the web, namely: i) cost: producing quality interoperable and linkable metadata is a costly and time-consuming process; ii) subjectivity: human annotators generally have specific views about content and the context in which it is used; iii) restrictiveness: a schema with few restrictions (such as free text fields) generally provides subjective and inconsistent terminology that is not easily machine readable; iv) longevity: constructing an annotation schema for specific purposes that is sufficiently generic to encompass different domains is a difficult task; v) privacy: metadata can contain private or confidential information, which requires special care; and vi) standardization: there is a need for syntactic and semantic-level standardization to achieve interoperability between different metadata schema and applications.

Based on these challenges, metadata as products and processes evidently require specific modeling aspects. Despite the comprehensive and commonly used definition of metadata found in the information science literature, its uses, syntaxes and applications differ in scale, complexity and cost. The results of this study contribute by providing possible solutions on how to efficiently index, catalog and retrieve multimedia content considering the numerous types of existing metadata for different needs and situations, as explained below.

The OMRM aimed to cover the functional characteristics established for multimedia metadata (content-independent, content-dependent and descriptive) aimed at representing these types of documents.

The *Media_Creation*, *Media_Classification*, *Media_Information* and *Media_Usage* classes are intended to ontologically organize metadata related to the management of information resources (content-independent – FR1), which can be applied to the realization of media and its content, thereby promoting the organization of high-level descriptions of different domains that require this semantic distinction to represent and retrieve information resources. Such classes were modeled based on a proposition of reuse of the Media Ontology (Lemos and Souza 2020), an ontology proposed to define a set of central annotation properties to describe multimedia content, along with a set of mappings between the main metadata formats currently in use, such as EXIF, IPTC, DIG35, Dublin Core, MPEG-7, among others. For example, the models used in Cultural Heritage Domains (e.g.: the Functional Requirements for Bibliographic Records - FRBR - Family of Models, or the

Library Reference Model - LRM) (IFLA 2009; Riva et al. 2017; Bekiari et al. 2024b) does not cover administrative metadata important for bibliographic universe (Riva et al. 2017, 15), such as copyright (covered by *Media_ Usage* class from OMRM model), preservation metadata or acquisition processes (covered by Media_Creation and Media_Information classes from OMRM model). We can also mention the need of semantically formalized (via axioms) information structures for different types of media (e.g.: a video collection, a music album, works of art, films, books), allowing content providers, such as digital libraries or museums, to identify and interpret the different copyrights associated with their multimedia content resources. In this sense, the aforementioned ontological classes of the OMRM model would allow, due to: i) the levels of extensibility based on the high-level DUL ontology; ii) the integration with complementary MPEG ISO/IEC[2] initiatives - such as the MPEG-21 Multimedia Framework; and iii) its set of standardized ontologies for the codification of media-related intellectual property rights information (Kudumakis et al. 2019); would allow for the transparent use of multimedia services across a wide range of networks and devices for diverse users.

The *visual-descriptor-parameter*, *localization-descriptor-parameter* and *LL_Audio_Descriptor* classes are aimed at the ontological organization of quantitative metadata (content-dependent – FR2) for visual aspects and localization in spatial, temporal and spatiotemporal regions, as well as audio in media content. For these types of metadata, the *Data Value* pattern of the DUL upper ontology, including relations between a *Quality*, *Region* and *Entity*, represent ambiguity-free features and data values for annotations of this nature. This paves the way, for example, for processing documents from a complex domain such as tangible cultural heritage (e.g.: archeological artifacts, sculptures, buildings), which typically involves the segmentation of physical or digitized structures. Thus, semantic annotation for digital measurements of the resulting 3D fragments (Catalano et al. 2020) creates reliable counterparts for the reassembly of the complete artifact. This favors the integrated search for and retrieval of heterogeneous federated data in the field of cultural heritage.

The *Navigation_Access* and *HL_Audio_Descriptor* classes (FR3) were proposed to ontologically and sequentially organize navigation and access to content personalized according to users' media preference (such as a movie synopsis), and high-level audio features, including a descriptor for indexing spoken content in audio streams (Martínez et al. 2002). Examples of applications include: i) a film or video recording in which a character speaks a particular word or sequence of words; the media support would then be recognized and the query would return content at the specific media position; ii) databases of spoken documents that enable the position of discourse in corresponding audio documents to be retrieved; and iii) retrieving a photograph annotated by a statement.

The classes related to the *Collection* and *Decomposition Patterns* (*Spatio_Temporal_Segment*, T*emporal_Segment* and *Spatial_Segment*) promoted the organization of multimedia collections and multimedia segments, respectively. Both are related to the *Annotation Pattern* which, in conjunction with its specialized classes of metadata types (described above), assign semantic-level links to and between media using LOD principles (Bizer et al. 2009) via specific annotation in content or in media realization (FR4).Structural annotations describe the structure of multimedia content in terms of video segments (frames, moving and static regions) and audio segments, while annotations on content describe real-world objects, agents, events and notions (FR3) that can be abstracted from multimedia content and linked to knowledge organization systems (Lemos and Souza 2020; Lemos et al. 2022) such as ontologies and SKOS of specific domains available online.

In the field of descriptive cataloging, the absence of semantic standards to describe multimedia digital objects at the levels of data structure, value, content and communication (Martins et al. 2022, 7) causes serious problems in standardization, normalization, quality and exchange of descriptions in a linked open data environment (Machado et al. 2019) that could be remedied by Semantic Web vocabularies that align with the by cataloging principles (Galeffi et al. 2016), for example the RDA Element Sets[3] and the RDA Value Vocabularies[4] that were created from attributes and relationships defined in Resource Description and Access (RDA[5]) at the RDA registry.

In this context, OMRM can contribute to the field of descriptive cataloging and its high-level principles focused on the user's tasks of finding, identifying, selecting, retrieving, navigating and exploring the item within an online catalog as a search and discovery system on the web. Therefore, the reference model seeks to portray generic and specific aspects of the multimedia document (at various levels of multimedia entities granularity) that make it unique by establishing well-founded semantic (with the use of a upper ontology) and interoperable access points that allow users greater search and retrieval capabilities in the web of data, making relationships explicit and providing contextual information.

Additionally, the OMRM can contribute to the FAIR guiding principles (Wilkinson et al. 2016) by allowing the creation of consistent and persistent identifiers (e.g.: URI and IRI, among others) of multimedia objects (data) and their annotations (metadata), whose formal semantics can be used in different online datasets for navigation, collection, extraction, mapping, enrichment, aggregation and other possible human or corresponding automated actions that favor the localization ('F'), access ('A'), interoperability

('I') and reuse ('R') functions. As reported by Guizzardi (2020), the "I" (interoperability) of FAIR is only possible with the support of information structures (e.g.: metadata standards, controlled vocabularies, cataloging rules and usage licenses) that are ontologically consistent and explain the ontological commitments that they make. According to the author, the description of real-world objects requires more than vocabularies, but the use and reuse of good domain ontologies.

As such, data providers that use a FAIR-based data quality policy, for example, could benefit from an aggregation service for online collections of multimedia objects (as a whole or in parts) to expand the search possibilities of their users in knowledge networks semantically linked to the web of data.

Problems associated with the syntactic and semantic interoperability requirements of online multimedia applications can be mitigated by the formal nature of the DUL upper ontology and its design patterns for ontology content. These structures ensure that the intended meaning of the captured semantics in the reference model can be shared between different applications within the scope of the Semantic Web.

5.0 Conclusion and future work

The results obtained in this study reflect numerous contributions to the fields of information and technology, particularly for knowledge and information organization and representation. Research in this area focuses largely on document processing (cataloging, indexing and classification), especially issues linked to the semantic nature of information. It should be noted that the present study used document processing concepts, theories, principles and methods, including content analysis, cataloging, classification, categorization and conceptual modeling. These practices supported the understanding, interpretation and systematization of knowledge resource content (metadata standards, controlled vocabularies, conceptual models and ontologies) in the OMRM architecture.

The central framework of the OMRM is a conceptual ontology based on cognitive, philosophical and linguistic aspects that provide metacategories to formally describe events, objects, time, space and others in order to semantically organize content from domain-specific ontologies. In this respect, formal semantics from the OWL representation language contribute considerably to the scope of the proposed conceptual structure that aims to describe any aspect related to multimedia data.

Comprehensiveness, in turn, is achieved via the application of Ontology Engineering principles, which suggests the use of upper ontologies and design patterns for ontology content. Thus, the OMRM seeks to ensure its connection with metadata, controlled vocabularies, SKOS and domain-specific ontologies via axiomatized definitions of high-level concepts from the DOLCE+DnS Ultralight ontology and its *Description and Situation*, *Information and Realization* and *Data Value* design patterns, used to generically organize entities associated with multimedia content, such as annotation, decomposition and collection. The central taxonomic structure of the proposed model therefore provides a solution to the first research question regarding *how to formally express a comprehensive conceptual framework that underlies the annotation domain for multimedia documents*.

The second question, *what methods and techniques would be suitable for selecting and aligning vocabularies and multimedia ontologies for the annotation of multimedia documents developed by different communities?*, was solved by using the NeOn Methodology guide, which is based on LOD initiatives for the construction of network ontologies. The methodology proved to be robust and efficient at explaining the different dimensions and variability in the analysis of knowledge resources identified in the literature and Semantic Web repositories, which ensured the reuse of ontologies suited to the conceptualization of the OMRM.

The third question, regarding *how to systematically organize existing types of metadata to annotate multimedia documents for different contexts and needs* was solved by covering functional requirements from metadata types (content-independent, content-dependent and descriptive) modeled here using relevant modeling reasoning that ensures the extensibility of the conceptualization to annotations (in both the realization of media and its content) aimed at different contexts and scenarios.

Thus, the objective of the study was achieved and contributes to the well-grounded proposal of an ontological conceptual model for the semantic organization of multimedia metadata for different application contexts that publish and consume data on the web.

It is important to note that the knowledge formalized in an ontology can benefit the user community that deals with multimedia documents in a number of ways. These include compiling intelligent queries with the possibility of expanding new concepts related to the initial query; helping to formulate information needs with automated reference services, including copyright information relating to digital media; supporting automated document annotation; and improving the user experience in semantic navigation between documents displayed on the web interface as search results.

Finally, it is important to underscore that the OMRM is, above all, a recommendation based on conceptual modeling principles involving multimedia ontologies and should, therefore, be implemented, coded, tested and validated in the field to obtain consistent conclusions on its applicability. We aim at applying the OMRM in studies on the integration of databases belonging to multimedia content providers with open semantic platforms (such as Wikidata), with a view to the creative collaboration of knowledge networks. This integration would make it possible to implement and test different federated queries with useful inference mechanisms based on previously established functional requirements, with a view to obtaining more conclusive results.

Notes

1. https://mpeg.chiariglione.org/standards/mpeg-7
2. https://mpeg.chiariglione.org/standards/exploration
3. https://www.rdaregistry.info/Elements/
4. https://www.rdaregistry.info/termList/
5. https://www.rdatoolkit.org/

References

Almeida, Maurício Barcellos and Fernanda Farinelli. 2017. "Ontologies for The Representation of Electronic Medical Records: The Obstetric and Neonatal Ontology". *Journal of the Association for Information Science and Technology* 68, no. 11: 2529-42. https://doi.org/10.1002/asi.23900

Andrews, Pierre, Ilya Zaihrayeu and Juan Pane. 2012. "A Classification of Semantic Annotation Systems". *Semantic Web* 3, no. 3: 223-48. https://doi.org/10.3233/SW-2011-0056

Bekiari, Chryssoula, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, Athanasios Velios, eds. 2024a. *Definition of the CIDOC Conceptual Reference Model*. Version 7.1.3. https://www.cidoc-crm.org/Version/version-7.1.3

Bekiari, Chryssoula, Martin Doerr, Patrick Le Bœuf, and Pat Riva. 2024b. *LRMoo object-oriented definition and mapping from the IFLA Library Reference Model*. https://www.cidoc-crm.org/lrmoo/ModelVersion/version-1.0-2

Bizer, Christian, Tom Heath and Tim Berners-Lee. 2009. "Linked Data - The Story So Far". *International journal on Semantic Web and information systems* 5, no. 3: 1–22. https://doi.org/10.4018/jswis.2009081901.

Borgo, Stefano and Claudio Masolo. 2009. "Foundational choices in DOLCE". In *Handbook on ontologies*, edited by S. Staab and R. Studer, 361-81. Berlin, Heidelberg: Springer Berlin Heidelberg.

Catalano, Chiara Eva, Valentina Vassallo, Sorin Hermon and Michela Spagnuolo. 2020. "Representing Quantitative Documentation of 3D Cultural Heritage Artifacts with CIDOC CRMdig". *International Journal on Digital Libraries* 21, no. 4: 359-73. https://doi.org/10.1007/s00799-020-00287-3

Charles, Valentine, Robina Clayphan, and Antoine Isaac, eds. 2017. *Definition of the Europeana Data Model v5.2.8*. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf

Dijkshoorn, Chris, Lizzy Jongma, Lora Aroyo, Jacco Van Ossenbruggen, Guus Schreiber, Wesley Ter Weele and Jan Wielemaker. 2018. "The Rijksmuseum Collection as Linked Data". *Semantic Web* 9, no. 2: 221-30. https://doi.org/doi: 10.3233/SW-170257

Europeana Tech. 2021. AI in relation to GLAMs Task Force: report and recommendations. https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/Europeana_Network_Task_Forces/Final_reports/AI%20in%20relation%20to%20GLAMs%20Task%20Force%20Report.pdf

Falbo, Ricardo de Almeida. 2014. "SABiO: Systematic Approach for Building Ontologies". In *Ontologies in Conceptual Modeling and Information Systems Engineering: Proceedings of the 1st Joint Workshop on Ontologies in Conceptual Modeling and Information Systems Engineering 21 September 2014,* edited by Giancarlo Guizzardi, Oscar Pastor, Yair Wand, Sergio de Cesare, Frederik Gailly, Mark Lycett and Chris Partridge. https://ceur-ws.org/Vol-1301/ontocomodise2014_2.pdf

Fernández-López, Mariano, Asunción Gómez-Pérez, and Natalia Juristo. 1997. "Methontology: from ontological art towards ontological engineering". In *AAAI-97 Spring Symposium Series: Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series 24-26 March 1997, Stanford, CA, USA*. 33-40. http://oa.upm.es/5484/

Ferrada, Sebastián, Nicolás Bravo, Benjamin Bustos, and Aidan Hogan. 2018. "Querying Wikimedia Images Using Wikidata facts". In *WWW '18: Companion Proceedings of the The Web Conference 23-27 April 2018 Lyon, France*, 1815-21. http://sferrada.com/pdf/wikiworkshop2018.pdf

Fink, Eleanor E. 2018. *American Art Collaborative (AAC): Linked Open Data (LOD) Initiative: Overview and Recommendations for Good Practices*. http://americanartcollaborative.org/wp-content/uploads/2018/03/AAC_LOD_Overview_Recommendations.pdf

Gangemi, Aldo and Valentina Presutti. 2009. "Ontology design patterns". In *Handbook on ontologies*, edited by S. Staab and R. Studer, 221-43. Berlin, Heidelberg: Springer Berlin Heidelberg.

Gilliland, Anne. J. 2016. "Setting the Stage". In *Introduction to Metadata,* edited by Murtha Baca. 3rd ed. Los Angeles: Getty Research Institute. http://www. getty.edu/publications/intrometadata/setting-the-stage/

Grüninger, Michael, and Mark S. Fox. 1995. *Methodology for the design and evaluation of ontologies*. http://www.eil.

utoronto.ca/wp-content/uploads/enterprise-modelling/papers/gruninger-ijcai95.pdf

Guizzardi, Giancarlo. 2020. "Ontology, Ontologies and the 'I' of FAIR". *Data Intelligence* 2, no. 1-2: 181-91. https://doi.org/10.1162/dint_a_00040

Higgins, Sarah. 2011. "Digital Curation: The Emergence of a New Discipline". *International Journal of Digital Curation* 6, no. 2: 78-88. https://doi.org/10.2218/ijdc.v6i2.191

IFLA (International Federation of Library Associations and Institutions). 2009. *Functional Requirements for Bibliographic Records: Final Report as Amended and Corrected Through February 2009*. https://repository.ifla.org/handle/20.500.14598/811

Galeffi, Agnese, María Violeta Bertolini, Robert L. Bothmann, Elena Escolano Rodríguez and Dorothy McGarry. 2016. *Statement of International Cataloguing Principles (ICP)*. Edition with minor revisions. Den Haag: IFLA. https://repository.ifla.org/handle/20.500.14598/80

Kudumakis, Panos, Thomas Wilmering, Mark Sandler, Víctor Rodríguez-Donce, Laurent Boch and Jaime Delgado. 2019. *MPEG Intellectual Property Rights Ontologies*. https://mpeg.chiariglione.org/standards/mpeg-21

Lancaster, Frederick Wilfrid. 1986. *Vocabulary control for information retrieval*. 2nd ed. Virginia: IRP.

Lemos, Daniela Lucas da Silva and Renato Rocha Souza. 2020. "Knowledge Organization Systems for the Representation of Multimedia Resources on the Web: A Comparative Analysis". *Knowledge Organization* 47, no. 4: 300-19. https://doi.org/10.5771/0943-7444-2020-4-300

Lemos, Daniela Lucas da Silva, Dalton Lopes Martins, Asla Medeiros e Sá, Luciana Conrado Martins and Danielle do Carmo. 2022. "A Proposal in Creating a Semantic Repository for Digital 3D Replicas: The Case of Modernist Sculptures in Public Spaces of Rio De Janeiro". *Knowledge Organization* 49, no. 3: 151-71. https://doi.org/10.5771/0943-7444-2022-3-151

Machado, Luis Miguel Oliveira, Renato Rocha Souza and Maria da Graça Simões. 2019. "Semantic Web or Web of Data? A Diachronic Study (1999 to 2017) of The Publications of Tim Berners-Lee and The World Wide Web Consortium". *Journal of the American Society for Information Science and Technology* 70, no. 7: 701–14. https://doi.org/10.1002/ asi.24111

Martínez, José M., Rob Koenen and Fernando Pereira. 2002. "MPEG-7: The Generic Multimedia Content Description Standard, Part 1". *IEEE multimedia* 9, no. 2: 78-87. https://doi.org/10.1109/93.998074

Martins, Dalton Lopes, Daniela Lucas da Silva Lemos, Luis Felipe Rosa de Oliveira, Joyce Siqueira, Danielle do Carmo and Vinicius Nunes Medeiros. 2022. "Information Organization and Representation in Digital Cultural Heritage in Brazil: Systematic Mapping of Information Infrastructure in Digital Collections for Data Science Applications". *Journal of the Association for Information Science and Technology* 73, no. 6: 1–20. https://doi.org/10.1002/asi.24650

Masolo, Claudio, Stefano Borgo, Aldo Gangemi, Nicola Guarino and Alessandro Oltramari. 2003. *Ontology Library (Wonder-Web Deliverable D18)*. http://www.loa.istc.cnr.it/old/Papers/D18.pdf

Messaoudi, Tommy, Philippe Véron, Gilles Halin and Livio De Luca. 2018. "An Ontological Model for the Reality-based 3D Annotation of Heritage Building Conservation State". *Journal of Cultural Heritage* 29: 100-12. https://doi.org/10.1016/j.culher.2017.05.017

Mora-Cantallops, Marçal, Salvador Sánchez-Alonso and Elena García-Barriocanal. 2019. "A Systematic Literature Review on Wikidata". *Data Technologies and Applications* 53, no. 3: 250-68. https://doi.org/10.1108/DTA-12-2018-0110

Nack, Frank, Jacco van Ossenbruggen and Lynda Hardman. 2005. "That Obscure Object of Desire: Multimedia Metadata on the Web, Part 2". *IEEE MultiMedia* 12, no.1: 54-63. https://doi.org/10.1109/MMUL.2005.12

Navarrete, Trilce, and Elena Villaespesa. 2021. "Image-based Information: Paintings in Wikipedia". *Journal of Documentation* 77, no.2: 359-80. https://doi.org/10.1108/JD-03-2020-0044

NISO (National Information Standards Organization). 2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005*. Baltimore: NISO.

Riva, Pat, Patrick LeBoeuf and Maja Žumer. 2017. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Robledano-Arillo, Jesús, Diego Navarro-Bonilla and Julio Cerdá-Díaz. 2020. "Application of Linked Open Data to The Coding and Dissemination of Spanish Civil War Photographic Archives." *Journal of Documentation* 76, no. 21: 67-95. https://doi.org/10.1108/JD-06-2019-0112

Shadbolt, Nigel, Tim Berners-Lee and Wendy Hall. 2006. "The Semantic Web Revisited". *IEEE intelligent systems* 21, no. 3: 96-101. https://doi.org/10.1109/MIS.2006.62

Saathoff, Carsten and Ansgar Scherp. 2010. "Unlocking the Semantics of Multimedia Presentations in the Web with the Multimedia Metadata Ontology." In *Proceedings of the 19th International Conference on World Wide Web, 26-30 April 2010 Raleigh NC, USA,* 831-40. New York: ACM. https://doi.org/10.1145/1772690.1772775

Silva, Daniela Lucas da, Renato Rocha Souza and Maurício Barcellos Almeida. 2012. "Ontologies and controlled vocabulary: comparison of building methodologies". In *Semantic Technologies for Business and Information Systems Engineering: Concepts and Applications,* edited by Stefan Smolnik, Frank Teuteberg and Oliver Thomas, 1-15. Hershey, PA: IGI Global.

Siqueira, Joyce and Dalton Lopes Martins. 2021. "Workflow Models For Aggregating Cultural Heritage Data on The Web: A Systematic Literature Review". *Journal of the Association for Information Science and Technology* 73, no. 2: 204–24. https://doi.org/10.1002/asi.24498

Suárez-Figueroa, Mari Carmen, Asunción Gómez-Pérez and Mariano Fernández-López, eds. 2012. "The NeOn Methodology for Ontology Engineering." In *Ontology Engineering in a Networked World,* edited by Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta and Aldo Gangemi, 9-34. Berlin: Springer, 9-34.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Cambridge: The MIT Press.

van Ossenbruggen, Jacco, Frank Nack and Lynda Hardman. 2004. "That Obscure Object of Desire: Multimedia Metadata on The Web, Part-1". *IEEE MultiMedia* 11, no.4: 38-48. https://doi.org/10.1109/MMUL.2004.36

Wilkinson, Mark, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niglas Blomberg *et al*. 2016. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific data* 3, no. 1. https://doi.org/10.1038/sdata.2016.18

Zeng, Marcia Lei and Jian Qin. 2016. *Metadata*. 2nd ed. Atlanta: ALA Neal-Schuman.