

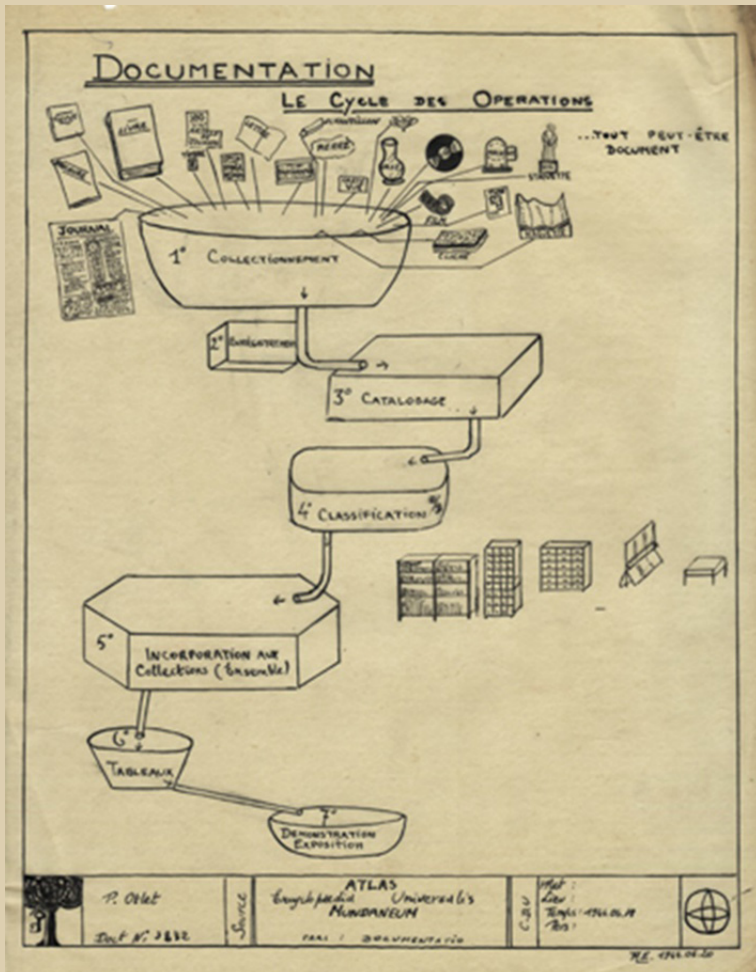
AIDa informazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

NUMERO 3-

ANNO 43

LUGLIO-DICEMBRE 2025



AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

Fondata nel 1983 da Paolo Bisogno

Proprietario della rivista:

Università della Calabria

Direttore Scientifico:

Roberto Guarasci, *Università della Calabria*

Direttore Responsabile:

Fabrizia Flavia Sernia

Comitato scientifico:

Anna Rovella, *Università della Calabria*;

Maria Guercio, *Sapienza Università di Roma*;

Giovanni Adamo, *Consiglio Nazionale delle Ricerche* †;

Claudio Gnoli, *Università degli Studi di Pavia*;

Ferruccio Diozzi, *Centro Italiano Ricerche Aerospaziali*;

Gino Roncaglia, *Università della Toscana*;

Laurence Favier, *Université Charles-de-Gaulle Lille 3*;

Madjid Ihadjadene, *Université Vincennes-Saint-Denis Paris 8*;

Maria Mirabelli, *Università della Calabria*;

Agustín Vivas Moreno, *Universidad de Extremadura*;

Douglas Tudhope, *University of South Wales*;

Christian Galinski, *International Information Centre for Terminology*;

Béatrice Daille, *Université de Nantes*;

Alexander Murzaku, *College of Saint Elizabeth, USA*;

Federico Valacchi, *Università di Macerata*.

Comitato di redazione:

Antonietta Folino, *Università della Calabria*;

Erika Pasceri, *Università della Calabria*;

Maria Taverniti, *Consiglio Nazionale delle Ricerche*;

Maria Teresa Chiaravallotti, *Consiglio Nazionale delle Ricerche*;

Assunta Caruso, *Università della Calabria*;

Claudia Lanza, *Università della Calabria*.

Segreteria di Redazione:

Valeria Rovella, *Università della Calabria*

Editrice: Cacucci Editore S.a.s.

Via D. Nicolai, 39 – 70122 Bari (BA)

www.cacuccieditore.it

e-mail: riviste@cacuccieditore.it

Telefono 080/5214220

AIDAinformazioni

RIVISTA SEMESTRALE DI SCIENZE DELL'INFORMAZIONE

«AIDAinformazioni» è una rivista scientifica che pubblica articoli inerenti alle Scienze dell'Informazione, alla Documentazione, all'Archivistica, alla Gestione Documentale e all'Organizzazione della Conoscenza ma amplia i suoi confini in ulteriori campi di ricerca affini quali la Terminologia, la Linguistica Computazionale, la Statistica Testuale, ecc. È stata fondata nel 1983 quale rivista ufficiale dell'Associazione Italiana di Documentazione Avanzata e nel febbraio 2014 è stata acquisita dal Laboratorio di Documentazione dell'Università della Calabria. La rivista si propone di promuovere studi interdisciplinari oltre che la cooperazione e il dialogo tra profili professionali aventi competenze diverse, ma interdipendenti. I contributi pubblicati affrontano questioni teoriche, metodologie adottate e risultati ottenuti in attività di ricerca o progettuali, definizione di approcci metodologici originali e innovativi, analisi dello stato dell'arte, ecc.

«AIDAinformazioni» è riconosciuta dall'ANVUR come rivista di Classe A per l'Area 11 – Gruppo Scientifico Disciplinare 11/HIST-04 – Scienze del libro, del documento e storico-religiose e come rivista scientifica per le Aree 10 – Scienze dell'antichità, filologico-letterarie e storico-artistiche; 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche; 12 – Scienze giuridiche; 14 – Scienze politiche e sociali. È anche annoverata dall'ARES (Agence d'évaluation de la recherche et de l'enseignement supérieur) tra le riviste scientifiche dell'ambito delle Scienze dell'Informazione e della Comunicazione. La rivista è, inoltre, indicizzata in: ACNP – Catalogo Italiano dei Periodici; BASE – Bielefeld Academic Search Engine; ERIH PLUS – European Reference Index for the Humanities and Social Sciences – EZB – Elektronische Zeitschriftenbibliothek – Universitätsbibliothek Regensburg; Gateway Bayern; KVK – Karlsruhe Virtual Catalog; The Library Catalog of Georgetown University; SBN – Italian union catalogue; Ulrich's; Union Catalog of Canada; LIBRIS – Union Catalogue of Swedish Libraries; Worldcat.

I contributi sono valutati seguendo il sistema del *double blind peer review*: gli articoli ricevuti sono inviati in forma anonima a due referee, selezionati sulla base della loro comprovata esperienza nei topics specifici del contributo in valutazione.

AIDAinformazioni

Anno 43

N. 3-4 – luglio-dicembre 2025

CACUCCI  EDITORE
BARI

PROPRIETÀ LETTERARIA RISERVATA

© 2025 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.it> e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

Sommario

GIOVANNA ARACRI, ANTONIETTA FOLINO, MARIA TERESA GUAGLIANONE, MARCO MANNA, Identification of thesaurus relationships with the support of ChatGPT and Gemini	9
ALESSANDRA BONSIGNORI, Terminologia giuridica e organizzazione della conoscenza	51
MARIA TERESA CHIARAVALLI, MARIA TERESA GUAGLIANONE, ELENA CARDILLO, Metadattazione delle tipologie di documenti clinici: un approccio alla standardizzazione mediante LOINC	73
CARLA BEATRIZ MARQUES FELIPE, CARLA MARIA MARTELLOTE VIOLA, RAIMUNDA FERNANDA DOS SANTOS, Domain Analysis on Data Management	91
CLAUDIO GNOLI, EMANUELA VALENZANO, Che faccetta sei?	111
MANUELA GRILLO, Strumenti di convergenza tra archivi e biblioteche: i thesauri	121
BIANCA LA MANNA, Rappresentare gli atti notarili medievali in Linked Open Data (LOD)	141
CLAUDIA LANZA, L'univocità e la variabilità terminologica a supporto della rappresentatività delle risorse semantiche di organizzazione e rappresentazione della conoscenza specialistica	163
LARISSA MORAES MARTINS, CARLOS EDUARDO DA SILVA CARVALHO, ANA CRISTINA DE ALBUQUERQUE, MARIO BARITÉ, Possibilities for using cultural warrant to complement literary work	181
BIANCA THEREZINHA CARVALHO PANISSET, LUCIA MARIA VELLOSO DE OLIVEIRA, LETICIA SOUZA DA COSTA SAMPAIO, CAMILLA CAMPOI DE SOBRAL, Identification of types of documents and knowledge organization: transparency and access in records management	197
SOYOUNG YOON, CLAUDIO GNOLI, ZIYOUNG PARK, Applying Integrative Levels Classification to Korean KOS Registry	211
EMANUELA NICOLE DONATO, GRAZIA SERRATORE, MARIA TAVERNITI, MARIA TERESA CHIARAVALLI, L'evoluzione della struttura documentale nelle cartelle cliniche psichiatriche dell'archivio storico del manicomio di Girifalco (1881-1978)	237
LORENZA TALARICO, Manuscripts in the Digital Age	263
MARIA LÍGIA TRIQUES, ANANDA FERNANDA DE JESUS, MARIO BARITÉ, JOSÉ EDUARDO SANTAREM SEGUNDO, ANA CRISTINA DE ALBUQUERQUE, Vocabularies in the digital context	275

MARCIN TRZMIELEWSKI, Context-Driven Use of Automatic Tools in Medical Ontology Design: A Case Study in Allergology	291
CARLA MARIA MARTELLOTE VIOLA, CARLA BEATRIZ MARQUES FELIPE, Exploring the relationships between the Cross-Cutting Women's Agenda - Multi-Year Plan 2024-2027 and the Knowledge Organization of legislative proposals under processing on women in Brazil	313
ELENA CENCETTI, CAMILLO CARLO PELLIZZARI DI SAN GIROLAMO, ELISABETTA VITI, Modellizzazione delle conoscenze: Thesaurus Nuovo soggettario e Wikidata, due strumenti a confronto	325
PONTISH YERAMYAN, ALEXANDER MURZAKU, NICK UVA, CURT ANDERSON, The transferability of brilliance	353
ANA FLÁVIA DIAS ZAMMATARO, ANA CRISTINA DE ALBUQUERQUE, GIULIA CRIPPA, LUCIA SARDO, CARLO BIANCHINI, MARIA LÍGIA TRIQUES, ISKO Italy in perspective	375

Contributi

Vocabularies in the digital context:

Terminological plurality in the Knowledge Organization

Maria Lígia Triques*, Ananda Fernanda de Jesus **, Mario Barité***,
José Eduardo Santarem Segundo****, Ana Cristina de Albuquerque*****

Abstract: The study investigated the approach to the term “vocabulary” in the digital context by the Knowledge Organization (KO) community. An exploratory and bibliographic analysis was conducted based on a research protocol. The analysis drew on studies that discussed the vocabularies used for assigning computational semantics within the Knowledge Organization KO journal (2004-2024). The results indicate that most studies concentrate on describing particular objects or domains, and the relationship with the digital context is frequently superficial or implied. The hypothesis of terminological plurality was confirmed (54 terms/expressions), with only 5 being repeated, and a notable absence of explicit definitions. The term “controlled vocabularies” was the most frequent, without differentiating between traditional and digital/computational. The absence of precise definitions and the generic use of the term represents a significant barrier to interdisciplinary research advancement. The context of use, structure and functionality are essential to clarify which concept of vocabulary is in focus.

Keywords: Controlled Vocabularies, Structured semantic vocabularies, Machine-Processable Vocabularies, Digital Environments, Knowledge Organization.

1. Introduction

Vocabularies have traditionally been the subject of research in library and information science (LIS), particularly in relation to Knowledge Organization (KO). In this area, vocabularies are a controlled instrument known as “controlled vocabularies”. These instruments are artificial relational languages that

* Universidade Estadual de Londrina – UEL, Londrina, Brasil. mlgia.triques@uel.br. ORCID: 0000-0002-0450-7138; albuati@uel.br. ORCID: 0000-0003-3506-0479..

** Universidade Estadual Paulista – Unesp, Câmpus de Marília, Brasil. af.jesus@unesp.br. ORCID: 0000-0001-7873-6040.

*** Universidad de la República – Udelar, Montevideo, Uruguay. mario.barite@fic.edu.uy. ORCID: 0000-0002-2992-6582.

**** Universidade Estadual Paulista – Unesp, Cmpus de Marília e Universidade de São Paulo – USP, Ribeirão Preto, Brasil. santarem@usp.br. ORCID: 0000-0003-3360-7872.

represent information and knowledge in a specific domain using a controlled and finite set of terms called descriptors (Kobashi 2008).

Studies on vocabularies have accompanied technological advances, especially those related to web development. These studies focus on tools in Linked Open Data (LOD) technologies, including value and metadata vocabularies (Zeng 2019). Vocabularies are also frequently referenced in discussions of Knowledge Organization Systems (KOS) in various contexts (Barité et al. 2024).

In this scenario, digital environments have become protagonists in the development of new ways of using and understanding vocabularies. These new approaches coexist with traditional notions that underpin this instrument, especially in its interdisciplinary nature with other fields, such as Linguistics, Lexicography, and Computing. These fields often act at the interface of Natural Language Processing (NLP), Artificial Intelligence (AI), and databases.

Technological possibilities introduce many variables into the creation and consumption of information. The dialogue about the use of vocabularies in digital contexts is influenced by various nomenclatures. Reaching a consensus in literature that allows for a precise definition is sometimes difficult. Various notions coexist and overlap with other terms, such as “set of metadata”, “ontologies” and “data/domain models”, to name a few. This makes it difficult to clearly identify and differentiate how these terms have been understood.

For the KO that works at the interface with traditional and emerging disciplines, this terminological imprecision makes it difficult to conduct interdisciplinary research that builds on assumptions related to vocabularies in analog and digital environments.

In this regard, we start from the premise that the KO literature offers few discussions that explicitly describe and differentiate the various perspectives on the use of “vocabulary” across interface areas. Although “vocabularies” is a basic term within KO, we argue that there is a plurality of terminologies shaped by the multiple uses introduced in the digital context. A context defined by the internet, digital technologies, and networked information systems.

To this end, a preliminary study was developed with the aim of identifying the definitions associated with the term “vocabularies” when applied to digital environments, more specifically in its relation to the attribution of computational semantics to data. The objective is to examine how the term “vocabularies,” when applied to digital contexts, has been addressed by the Knowledge Organization (KO) scientific community. Furthermore, the study seeks to confirm the hypothesis that there is a terminological plurality regarding the use of the term vocabulary in digital contexts.

Therefore, the purpose of this study is not to present an exhaustive analysis but rather to identify the terms that have been employed in relation to this theme. To achieve this, we conducted an initial exploratory study using

the journal *Knowledge Organization (KO)* as a reference, given its significant scientific output in the field and its status as a privileged venue for observing conceptual, terminological, and methodological trends. This choice makes it possible to map, although partially, how specialized literature has employed and consolidated different terms around the theme, providing a basis for future, more in-depth investigations.

In this sense, the proposal is justified so far as demonstrating the existing terminological plurality and systematizing the terms in use constitutes a fundamental step toward supporting and guiding subsequent research that investigates deeper into the discussions on such emerging usages, in coexistence with traditional ones, and identifying their correlations and distinctions.

2. Vocabularies: organizational structures

In a broad sense, the term “vocabulary” refers to a set of words belonging to a particular language or area of knowledge. In the specific context of areas such as Information Science that dialogue with KO, vocabularies take on a structured form, involving the selection and standardization of terms to accurately describe concepts, allowing for the standardized representation of concepts present in documents, objects or data (Zeng and Chan 2004).

KO has focused on the challenges and opportunities arising from the emerging contexts associated with information environments. These environments, especially digital ones, are characterized by a diversity of media, formats and data flows, requiring methods and instruments capable of guaranteeing the representation, retrieval and meaningful circulation of information. Historically associated with KO tools such as subject heading lists, classifications and thesauruses, vocabularies are organizational structures that accompany technological and socio-cultural transformations, reflecting the way they are conceived, implemented and understood.

The advent of the Semantic Web and the growing demand for interoperability between heterogeneous systems has led to the development of ontologies, which are formalized conceptual schemes and machine-processable models. These ontologies enable a richer and more connected representation of knowledge (Zeng 2008).

A brief review of the vocabularies evolution as organizational structures shows a trajectory marked by transformations that accompany the ways in which information is represented, structured and retrieved over time. These tools can be thought of from their consolidation as linguistic control mechanisms to their configuration as interoperable semantic structures in the digital sphere.

As Barité (2011) explains, although there are different understandings of what constitutes these instruments, such as the North American NISO stan-

dard and the ALA Glossary, there are broader conceptions that include all languages used to represent and organize knowledge as controlled vocabularies under a conceptual umbrella.

Vocabularies are therefore often understood within the broader scope of KOS, an expression that covers a diverse set of tools aimed at the structured representation of knowledge (Barité 2011). KOS are instruments that enable the organized interpretation of knowledge domains, contributing to terminological standardization, conceptual navigation and interoperability between information systems (Zeng and Chan 2004).

Zeng and Chan (2004, 1) highlight three fundamental categories among the main types of KOS based on characteristics such as structure and complexity, relationships between terms and historical function as defined in 2000 by NKOS (Zeng and Chan 2004). These categories are:

1. Term lists, which contain lists of words or phrases, often with definitions. Examples include authority files, glossaries, gazetteers, and dictionaries.

2. Classifications and categorization schemes, which emphasize the creation of subject sets. The most notable examples are library classification schemes, taxonomies, and categorization schemes.

3. Relational vocabularies, which emphasize the relationships and connections between terms and the concepts they represent, including lists of subject headings, thesauri, semantic networks, and ontologies.

Thus, the adoption of machine-processable vocabularies can be understood as an extension and reconfiguration of traditional knowledge representation practices, now mediated by specific languages and standards, such as the Resource Description Framework (RDF), the Web Ontology Language (OWL) and the Simple Knowledge Organization System (SKOS). These standards enable the formalization and interoperability of vocabularies in distributed digital environments, allowing not only the standardized description of concepts, but also the creation of semantic links between data from different sources. This contributes to the construction of more integrated and accessible knowledge networks and enables automatic inferences of relationships and links with other networked datasets (Zeng 2008; Jia 2020).

In this context, vocabularies provide the formal semantics of the concepts and relationships used to describe data and represent an area of interest, which makes them the foundation for Linked Data and Linked Open Data (Jia 2020). As Bizer, Heath and Berners-Lee (2009) discuss, while Linked Data is an approach based on best practices for making structured data interconnected, interoperable and accessible on the Web, using Semantic Web technologies, Linked Open Data (LOD) is the subset of Linked Data that adds the dimension of openness, i.e. the data is made freely available for use, reuse and redistribution.

In the Semantic Web, vocabularies then have their functions expanded to the various possible types of applications, which include classifying terms that can be used in a specific context, characterizing relationships and defining restrictions imposed on the use of terminologies (World Wide Web Consortium 2013).

As Jia (2020) explains, there are therefore various types of vocabulary that are generally recognized based on their structural and functional complexity and focus on fulfilling primary functions, which more complex structures generally perform most or all. Such functions can include the elimination of ambiguities, indexing and cataloging, as well as query expansion and the integration of information using identification, description and control of terms and conceptual relationships (Jia 2020).

The study by Jia (2020) proposes a comprehensive analysis of the categorization of semantic vocabularies, highlighting that they can be grouped according to different perspectives, depending on their function, structure and mode of availability. The first perspective considers RDF graph standards. In this approach, vocabularies are divided into sets of metadata elements, which define classes and attributes needed to describe resources such as people, organizations or places; and value vocabularies, such as thesauruses and subject classification systems, which provide controlled schemes for standardizing the values assigned to elements, ensuring uniformity and interoperability in data representation.

The second perspective refers to the degree of formalization. In this case, vocabularies are described using formal Semantic Web languages, such as RDF, which structures relationships between entities, OWL, which allows the definition of restrictions and complex logical inferences, SKOS, used to represent basic structures of controlled vocabularies, such as thesauruses, and RIF (Rule Interchange Format), which enables the exchange of rules between heterogeneous applications (Jia 2020). The third perspective proposed by Jia (2020) concerns the vocabularies openness. From this point of view, they can be classified as: vocabularies of limited openness, which make only part of their content available through licensing or restrictions on use; and fully open vocabularies, which offer unrestricted access through interfaces such as public APIs, favoring interoperability and the reuse of data by various applications.

In this sense, vocabularies range from linguistic control tools applied to indexing in analog environments to becoming elements of semantic infrastructure on the Web. The evolution of these systems is concomitant with the technological, conceptual, and methodological transformations of KO, reflecting the demands for interoperability, openness, granularity, and contextualized representation of information.

3. Methodological Procedures

An exploratory and bibliographical analysis is conducted in this study in order to identify how the term “vocabulary”, when applied to digital contexts, has been approached by the scientific community of knowledge organization.

To this end, this study provides, from an initial and exploratory survey of the proposed topic, a mapping and alignment of terminological plurality in the KO that permeates the concept of structured vocabularies in a machine-processable format. The objective of this study is not to provide an exhaustive analysis, but rather to identify the terms that have been employed to refer to this subject.

To capture this plurality, a research protocol (Tab. 1) was applied, based on the research question: *How has Knowledge Organization named and defined the term vocabularies when referring to structured vocabularies in machine-processable format, developed and applied with the aim of promoting semantic representation in digital contexts?*

Research question (main)	How has the Knowledge Organization community named and defined the term vocabularies, when referring to structured vocabularies in machine-processable format, developed and applied with the aim of promoting semantic representation in digital contexts?
Objectives	Identify the terms and definitions used by the Knowledge Organization community when referring to structured vocabularies in machine-processable format, developed and applied with the aim of promoting semantic representation in digital contexts
Search strategy	(“semantic” OR “data”) AND “vocabulary”; “vocabulary”
Databases consulted	Search in title and abstracts, applying the Source Title: KNOWLEDGE ORGANIZATION and Fields of Research (ANZSRC 2020) filters: 4610 Library and Information Studies; Source Title: KNOWLEDGE ORGANIZATION
Period covered	2004-2024
Idioms	English

Inclusion Criteria	(I) Discusses the use of applied vocabularies in establishing computational semantics in digital contexts
Exclusion Criteria	(E) Not in the languages established for the search; (E) Not in the period covered by the research (E) Only mentions the topic, without going into depth or presenting relevant definitions; (E) Does not address the topic of interest; (E) It was not possible to obtain access to the complete document; (E) Addresses vocabularies aimed at the analogical context (E) Addresses the application of traditional
Extraction form	1) Term used; 2) Definition; 3) Focus of the document.
Date of collection	21/11/2024

Table 1. Research Protocol Adopted.

The choice of both the database used for the Exploratory Analysis and the filters applied is related to the focus of the study of specific terminology to be carried out later, since its focus will be the analysis of the term in the context of knowledge organization. It was therefore assumed that the ISKO, in its international chapter, represents a relevant and sufficiently representative cross-section to make up the exploratory analysis, and that the Knowledge Organization KO journal, in addition to presenting the work presented at the international ISKO:

[...] is a forum for all those interested in the organization of knowledge on a universal or a domain-specific scale, using concept-analytical or concept-synthetic approaches, as well as quantitative and qualitative methodologies. KNOWLEDGE ORGANIZATION also addresses the intellectual and automatic compilation and use of classification systems and thesauri in all fields of knowledge, with special attention being given to the problems of terminology (Knowledge Organization 2025).

The time frame, which covers the last two decades, is justified by the emergence of the topic alongside the development of the World Wide Web. However, in the last two decades, approaches within and outside the digital sphere have coexisted and complemented each other, often leaving an implicit distinction between when the concept of structured vocabularies in computational semantics is or is not used in the study, suggesting a lack of conceptual rigor. Consequently, it was deemed necessary to employ exclusion criteria in instances where the study does not pertain to this specific context.

4. Results and Discussions

Based on the strategy presented, a total of 55 documents were retrieved, which underwent a selection process that identified 1 duplicate document, 36 rejected documents and 18 accepted documents to make up this preliminary research *corpus* (Tab. 2) shows the list of documents that make up the research *corpus* and their respective thematic focuses.

Id	Authors	Title	Focus
02	Baca e Gill (2015)	Encoding Multilingual Knowledge Systems in the Digital Age: the Getty Vocabularies	Present the history and evolution of the electronic thesauruses produced and maintained by the Getty Research Institute (GRI).
10	Chen (2019)	Semantic Enrichment of Linked Archival Materials	Enhanced a semantic data model for Chen Cheng-po's archival materials of paintings and documents.
18	Dobreski, Qine e Resnick (2020)	Depicting Historical Persons and Identities: A Faceted Approach	Developed a seven-facet structure of personal identity, based on a collection of 19th and 20th century images, aiming to contextualize terms and historical perspectives, improving user understanding.
27	Graf (2020)	Domain Analysis Applied to Online Graffiti Art Image Galleries to Reveal Knowledge Organization Structures Used Within an Outsider Art Community	Addresses a methodology based on domain analysis for the elucidation of knowledge organization systems used in image websites documenting graffiti, graffiti art and street art.
36	Lemos e Souza (2020)	Knowledge Organization Systems for the Representation of Multimedia Resources on the Web: A Comparative Analysis	Presents and compares existing metadata standards, vocabularies and ontologies for describing multimedia resources.
40	Marcondes (2020)	Towards a Vocabulary to Implement Culturally Relevant Relationships Between Digital Collections in Heritage Institutions	Discusses the development of Linked Open Data vocabulary to characterize the culturally relevant relationships of the collections of cultural heritage institutions.

47	Smith (2020)	Physics Subject Headings (PhySH)	Discusses the PhySH (Physics Subject Headings) vocabulary, a controlled faceted hierarchical vocabulary for physics, with some basic terms from related areas
12	Coladangelo (2021)	Organizing Controversy: Toward Cultural Hospitality in Controlled Vocabularies Through Semantic Annotation	Addresses the development of a controlled vocabulary to represent dance concepts for country dance choreography, focusing on discussing the implications of cultural and ethical issues related to the representation of gender and race.
30	Smith (2021)	RILM Index (Répertoire International de Littérature Musicale)	Discusses the creation of a thesaurus based on the Répertoire International de Littérature Musicale (RILM) Index, a partially controlled vocabulary designed to index academic writings on music and related subjects.
50	Velios e St.John (2021)	Linked Conservation Data: the Adoption and Use of Vocabularies in the Field of Heritage Conservation for Publishing Conservation Records as Linked Data	Describes an approach to improve the landscape of vocabularies in the field of conservation, providing guidelines for codification and alignment of vocabularies for future implementation of Linked Data.
14	Lemos et al. (2022)	A Proposal in Creating a Semantic Repository for Digital 3D Replicas: The Case of Modernist Sculptures in Public Spaces of Rio De Janeiro	It presents the results of the analysis of vocabularies for documents of physical artifacts and their digital counterparts on the Semantic Web.
26	Grabus et al. (2022)	Temporal Concept Drift and Alignment: An Empirical Approach to Comparing Knowledge Organization Systems Over Time	It presents a methodological contribution on how to approach changes in knowledge organization systems over time, improving the contextualization of historical humanities resources.

39	Marcondes (2022)	The Role of Vocabularies in the Age of Data: The Question of Research Data	Discusses how vocabularies can contribute to assigning computational semantics to digital research data in the context of Big Data.
48	Smith (2022)	Simple Knowledge Organization System (SKOS)	It discusses the Simple Knowledge Organization System (SKOS), a Linked Data vocabulary for representing knowledge organization systems.
05	Barité, Parentelli e Rauch (2024)	Knowledge Organization Systems in Reference Works Specialized in Knowledge Organization: A Study of Punctual Terminology	Identifies records of the term 'knowledge organization system' and its synonymous or similar expressions in dictionaries, glossaries, vocabularies and encyclopaedias specializing in knowledge organization.
08	Binding e Tudhope (2024)	KOS-based Enrichment of Archaeological Fieldwork Reports	It discusses automatic indexing techniques (for final intellectual judgment) based on knowledge organization systems.
22	Georgaki (2024)	A Maritime Heritage Thesaurus Based on a Greek Project Documentation Case	It proposes a thesaurus in the field of Maritime Heritage with the aim of improving methods of archiving oral sources, written accounts, images, 3D archaeology and other multimedia objects.
23	Gnoli et al. (2024)	Library Catalog's Search Interface: Making the Most of Subject Metadata	It addresses the underuse of elements of knowledge organization systems in online library catalogs, hindering effective search and discovery by subject.

Table 2. List of accepted documents and their respective approaches.

Although the last 20 years have been established as a time frame, discussions about vocabularies in the context discussed by the research are relatively recent in the Knowledge Organization scientific community and have only been explored in the last 10 years, with a significant increase in interest in the subject since 2020.

Most studies focus on presenting or discussing the construction of vocabularies for describing specific objects such as paintings, images, 3D works, graffiti and street art. They also focus on describing specific domains such as naval, cultural heritage and music for academic contexts. Although these studies deal with bringing vocabularies closer to the digital context through the proposed applications, this is not the main focus of the texts, and the issue is only mentioned in a superficial or subtextual way.

Among the accepted articles, Marcondes (2022) stands out for the purposes of this study, as he discusses the importance of vocabularies for assigning computational semantics to digital data. The article points out that when adapted to the context of Linked Data, vocabularies make it possible to express the semantics of data, both for humans and machines, facilitating processing by computer agents.

After identifying and analyzing the documents that make up the research *corpus*, we tried to identify the terms adopted by the authors to refer to vocabularies in the context analyzed. In some cases, authors use expressions to refer to the concept; these expressions were also collected.

The terms and expressions identified, along with the number of studies in which these terms appear, are presented (Tab. 3). A single text may utilize multiple terms or expressions.

Terms and Expressions	Occurrence
controlled vocabulary/controlled vocabularies	10
metadata vocabulary/vocabularies	4
value vocabulary	2
semantic vocabularies	2
KOS [knowledge organization system] vocabularies	2
vocabulary/vocabularies	2
semantically structured vocabularies	1
vocabularies as linked data	1
terminological enrichment of controlled vocabulary tools	1
data vocabularies	1
vocabularies published as LOD [linked open data]	1
data vocabularies enrichment	1
standardized vocabularies	1
formalized vocabulary	1
vocabularies (or ontologies)	1
KOS [knowledge organization system] terms	1
vocabularies of different types and dimensions	1
controlled vocabularies through semantic annotation	1
controlled vocabularies to enrich the data	1

KOS[knowledge organization system] entry vocabularies	1
specialized vocabularies	1
languages and vocabularies based on standards recommended by the World Wide Web Consortium (W3C)	1
structured vocabulary	1
LOD [linked open data] vocabularies	1
controlled vocabularies for semantic enrichment	1
LOD KOS [linked open data knowledge organization system]	1
vocabularies are available as structured data in a format which can be machine-processed	1
machine readable vocabularies	1
vocabularies expressed as semantic web technologies	1
vocabulary (concepts, terms)	1
vocabularies of reference	1
vocabulary in annotating digital heritage objects published as LOD [linked open data]	1
controlled/standardized vocabularies	1
RDF vocabularies	1
controlled vocabularies on the web	1
web of data context vocabularies	1
controlled languages	1
ontology vocabularies	1
Total	54

Table 3. List of terms/expressions.

As can be seen in Table 3, there is a wide range of terminology in the OC used to refer to the concept of vocabularies in the context studied. A total of 54 terms/expressions were identified, of which only 6 are repeated in more than one text. The other terms/expressions occur in only one of the texts analyzed.

The term with the highest occurrence is “controlled vocabularies”, which does not allow a distinction to be made between traditional controlled vocabularies and those created to refer to structured vocabularies in a machine-processable format, developed and applied with the aim of promoting semantic representation in digital contexts.

Among the single occurrence terms, there are a number of variations that mention the context of Linked Data, the semantic web and the web of data, terms such as: “semantic vocabulary” (2); “LOD vocabularies” (1); “vocabularies published as LOD” (1); “vocabularies expressed as semantic web technologies” (1); “Vocabularies as Linked Data” (1); “semantically structured vocabularies” (1); “languages and vocabularies based on standards recommended by

er, this definition only becomes explicit through semantic associations with terms such as “semantic,” “LOD,” “expressed as semantic web technologies,” and “web of data context vocabularies”. The analysis reveals a limited presence of explicit definitions for the concept of vocabulary, with only five instances clearly identified.

The first (ID-02) defines vocabularies, in the context of the Getty Vocabularies, as a central data structure mapped to a common schema and interconnected both technically and semantically. This definition also describes the term “value vocabulary” within the Semantic Web domain as patterns of semantically structured, machine-readable data values, such as authority files, thesauri, subject headings, and controlled lists, used to populate metadata elements. The second definition (ID-39) characterizes “vocabularies” as sets of descriptive metadata representing specific entities or resources in the Web context. In this sense, value vocabularies assign standardized data values to descriptive items associated with entity instances described by metadata vocabularies.

Another definition (ID-05) refers to the expression “LOD KOS,” citing Soergel (2009), as a broad term encompassing all value vocabularies and ontology levels within the Semantic Web framework. It further argues that KOS are also used by computer programs to reason over data and must represent formal knowledge about concepts. The fourth definition (ID-08) addresses vocabularies in the context of semantic enrichment techniques and tools based on KOS, applied to support information discovery, as exemplified by the Europeana experience. Finally, the fifth definition (ID-23) understands the term “controlled vocabulary,” in the context of KOS, as any controlled list of terms used in metadata to describe documents.

The low incidence of definitions makes it difficult to understand whether the studies really refer to the context of computational semantics. In most cases, it is through clarifications regarding structure and functionality that the intended meaning can be grasped. However, the distinction between terms such as “ontology vocabularies,” “data vocabularies,” “metadata vocabularies,” and “value vocabularies” is not made explicit in most cases.

5. Conclusions

The aim of this study was to identify how the term “vocabulary”, when applied to digital contexts, specifically in reference to structured vocabularies in machine-processable formats, has been approached by the scientific community of knowledge organization.

Based on the *corpus* constructed, it was possible to observe that most of the studies focus on the development and discussion of vocabularies for the description of specific information objects, as well as the description of different domains such as naval, cultural heritage and music for academic contexts.

No studies were identified that specifically aimed to discuss “vocabularies” in machine-processable formats, nor that examined the impact of this context on the structure and functioning of these tools; the topic was mentioned only superficially or implicitly. The study confirms the hypothesis that in the literature on KO there are few discussions that recognize and differentiate the various understandings of the uses of the term vocabulary.

The identification and analysis of the terms used by the authors to refer to vocabularies created for the digital context confirmed the hypothesis that there is a wide terminological plurality in relation to the term studied, with 39 variations observed in just 18 articles analyzed. In addition to the plurality, there was a lack of definitions for the term, making the scenario even more complex, especially in interdisciplinary studies with approaches related to Information Science and Computer Science.

The very generic use of the term carries conceptual imprecision, in that it is often used as a synonym for the most varied types of KOS, which in essence have differences. In addition, new KOS approaches are outlined in computational semantics, which are referred to simply as “vocabularies”. These encompass efforts undertaken within the Semantic Web framework, without a precise differentiation.

In this way, the context of use in terms of structure and functionality is what guarantees the explicitness of which vocabulary concept is in focus. However, the lack of an explicit conceptual definition demonstrates the emergence of this relational field and the need for a study of specific terminology in which vocabularies play a leading role.

In this sense, we highlight the need for clarity and in-depth study of the terminology and conceptualization of the term “vocabulary”. This domain is characterized by plurality and interdisciplinarity, and the lack of conceptual clarity can act as a substantial impediment to progress in teaching and research related to the subject.

As future studies, we hope to expand the *corpus* of studies, adding other databases related to Knowledge Organization. Based on this expansion, the next stage is to carry out a study of specific terminology, the aim of which is to promote greater clarity and in-depth study of the various understandings related to the application and structuring of vocabulary in the digital context. This includes adding discussions on overviews and definitions of other terms, such as ontologies and topic maps, in order to encompass this terminological universe.

Acknowledgments

We acknowledge the support of national funding agencies: *Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)* and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*.

References

- Barité, Mario. 2011. "Sistemas de organización del conocimiento: Una tipología actualizada." *Informação & Informação* 16 (3): 122–39. <https://ojs.uel.br/revistas/uel/index.php/informacao/article/download/9952/pdf/0>.
- Barité, Mario, Varenka Parentelli, and Mirtha Rauch. 2024. "Knowledge Organization Systems in Reference Works Specialized in Knowledge Organization: A Study of Punctual Terminology." *Knowledge Organization* 51 (5): 340–61. <https://doi.org/10.5771/0943-7444-2024-5-355>.
- Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data – The Story So Far." *International Journal on Semantic Web and Information Systems* 5 (3): 1–22. <https://doi.org/10.4018/jswis.2009081901>.
- Jia, Junzhi. 2020. "From Data to Knowledge: The Relationships Between Vocabularies, Linked Data and Knowledge Graphs." *Journal of Documentation* 77 (1): 93–105. <https://doi.org/10.1108/JD-03-2020-0036>.
- Knowledge Organization*. 2025. IMR Press. Accessed July 11, 2025. <https://www.imrpress.com/journal/KO>.
- Kobashi, Nair Yumiko. 2008. *Vocabulário controlado: Estrutura e utilização*. Brasília: Escola Nacional de Administração Pública. <https://repositorio.enap.gov.br/bitstream/1/1289/41/Vocabul%C3%A1rio%20controlado%20-%20estrutura%20e%20utiliza%C3%A7%C3%A3o.pdf>.
- Marcondes, Carlos H. 2023. "The Role of Vocabularies in the Age of Data: The Question of Research Data." *Knowledge Organization* 49 (7): 467–82. <https://repositorio.ufmg.br/handle/1843/52263>.
- Soergel, Dagobert. 2009. *Knowledge Organization Systems: Overview*, 1-44.
- World Wide Web Consortium (W3C). 2013. *Linked Data Glossary*. Accessed July 11, 2025. <https://www.w3.org/TR/ld-glossary/>.
- Zeng, Marcia Lei. 2008. "Knowledge Organization Systems (KOS)." *Knowledge Organization* 35 (2–3): 160–82. <http://dx.doi.org/10.5771/0943-7444-2008-2-3-160>.
- Zeng, Marcia Lei. 2019. "Interoperability." *Knowledge Organization* 46 (2): 122–46. <https://www.nomos-elibrary.de/10.5771/0943-7444-2019-2-122.pdf>.
- Zeng, Marcia Lei, and Lois Mai Chan. 2004. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology* 55 (5): 377–95. <https://doi.org/10.1002/asi.10387>.

AID Ainformazioni

Rivista semestrale di Scienze dell'Informazione

Anno 43

N. 3-4 – luglio-dicembre 2025

Contributi

GIOVANNA ARACRI, ANTONIETTA FOLINO, MARIA TERESA GUAGLIANONE, MARCO MANNA, Identification of thesaurus relationships with the support of ChatGPT and Gemini

ALESSANDRA BONSIGNORI, Terminologia giuridica e organizzazione della conoscenza

MARIA TERESA CHIARAVALLOTI, MARIA TERESA GUAGLIANONE, ELENA CARDILLO, Metadattazione delle tipologie di documenti clinici: un approccio alla standardizzazione mediante LOINC

CARLA BEATRIZ MARQUES FELIPE, CARLA MARIA MARTELOTE VIOLA, RAIMUNDA FERNANDA DOS SANTOS, Domain Analysis on Data Management

CLAUDIO GNOLI, EMANUELA VALENZANO, Che faccetta sei?

MANUELA GRILLO, Strumenti di convergenza tra archivi e biblioteche: i thesauri

BIANCA LA MANNA, Rappresentare gli atti notarili medievali in Linked Open Data (LOD)

CLAUDIA LANZA, L'univocità e la variabilità terminologica a supporto della rappresentatività delle risorse semantiche di organizzazione e rappresentazione della conoscenza specialistica

LARISSA MORAES MARTINS, CARLOS EDUARDO DA SILVA CARVALHO, ANA CRISTINA DE ALBUQUERQUE, MARIO BARITÉ, Possibilities for using cultural warrant to complement literary work

BIANCA THEREZINHA CARVALHO PANISSET, LUCIA MARIA VELLOSO DE OLIVEIRA, LETICIA SOUZA DA COSTA SAMPAIO, CAMILLA CAMPOI DE SOBRAL, Identification of types of documents and knowledge organization: transparency and access in records management

SOYOUNG YOON, CLAUDIO GNOLI, ZIYOUNG PARK, Applying Integrative Levels Classification to Korean KOS Registry

EMANUELA NICOLE DONATO, GRAZIA SERRATORE, MARIA TAVERNITI, MARIA TERESA CHIARAVALLOTI, L'evoluzione della struttura documentale nelle cartelle cliniche psichiatriche dell'archivio storico del manicomio di Girifalco (1881-1978)

LORENZA TALARICO, Manuscripts in the Digital Age

MARIA LÍGIA TRIQUES, ANANDA FERNANDA DE JESUS, MARIO BARITÉ, JOSÉ EDUARDO SANTAREM SEGUNDO, ANA CRISTINA DE ALBUQUERQUE, Vocabularies in the digital context

MARCIN TRZMIELEWSKI, Context-Driven Use of Automatic Tools in Medical Ontology Design: A Case Study in Allergology

CARLA MARIA MARTELOTE VIOLA, CARLA BEATRIZ MARQUES FELIPE, Exploring the relationships between the Cross-Cutting Women's Agenda - Multi-Year Plan 2024-2027 and the Knowledge Organization of legislative proposals under processing on women in Brazil

ELENA CENCETTI, CAMILLO CARLO PELLIZZARI DI SAN GIROLAMO, ELISABETTA VITI, Modellizzazione delle conoscenze: Thesaurus Nuovo sogettario e Wikidata, due strumenti a confronto

PONTISH YERAMYAN, ALEXANDER MURZAKU, NICK UVA, CURT ANDERSON, The transferability of brilliance

ANA FLÁVIA DIAS ZAMMATARO, ANA CRISTINA DE ALBUQUERQUE, GIULIA CRIPPA, LUCIA SARDO, CARLO BIANCHINI, MARIA LÍGIA TRIQUES, ISKO Italy in perspective



mundaneum

In copertina

Disegno di Paul Otlet, Collections Mundaneum, centre d'Archives, Mons (Belgique).

ISBN 979-12-5965-639-1



9 791259 656391

ISSN 1121-0095



9 770112 100950