Metadata Standards: a Review Towards Modeling Experiments

Filipe A. Santana, André F. R. Cordeiro, Edson OliveiraJr

[1]Departamento de Informática – Universidade Estadual de Maringá (UEM)
Maringá – PR – Brazil

ra123515@uem.br, cordeiroandrefelipe@gmail.com, edson@din.uem.br

Abstract. *Metadata consists of elements that describe specific content of data. When structured according to standards, they contribute to data interoperability. This study presents a secondary study, a Systematic Mapping (SM), to search for metadata patterns that can be used in the field of Computer Science (CS), especially in the subjects of Software Engineering (SE) and Information Systems (IS). At the end of the search string execution, 43 studies were analyzed. Of these, 23 studies were selected. Based on the observed results, a promising research area is identified, focusing on the adoption of metadata standards in various activities within the context of CS, SE, and IS.*

1. Introduction

Metadata consists of elements that describe details about data, such as content, format, purpose, and structure. This description can enhance data retrieval and management [Al-Khalifa and Davis 2006]. When metadata is established and organized according to standards, it is observed that information exchange between different systems becomes more efficient [NISO 2017].

In this study, a secondary study, specifically a Systematic Mapping (SM), was conducted. To develop the SM, other secondary studies were consulted [Tauil 2018]; [Cerrao 2019]; [da Silva 2020], and [James 2020]. The SM was structured based on a protocol and executed with the aim of identifying metadata patterns that could be used in one or more research fields within Computer Science, with particular attention to the fields of Software Engineering (SE) and Information Systems (IS). Over 40 studies were analyzed during the course of the SM. Among the generated results are a list of identified metadata standards, fields, and subfields of research in CS that consider the use of the mentioned standards, as well as a list of related works. The following sections provide details of the conducted SM. Section 2 presents the Theoretical Foundation. The Research Methodology is presented in Section 3. Results are outlined in Section 4. Discussions of the Results, Threats to Validity, and Final Considerations are presented in sections 5, 6, and 7, respectively.

2. Research Studies about Metadata

Metadata can be defined as data about data [Hayslett 2023]. Metadata describes the content, format, purpose, and structure of data [Al-Khalifa and Davis 2006]. Moreover, it consists of elements that aid in data management and contribute to data retrieval. According to [Kaur et al. 2023], ”metadata can be used to determine where data is generated,

stored, and accessed, reflecting the specialization of metadata in recognizing and locating materials relevant to the user."

The use of metadata can be considered essential for the performance of a system [Gao et al. 2019]. Metadata is utilized in various areas and subfields of science [Formenton et al. 2018]. An example is the use of metadata in Open Data initiatives for the publication of large volumes of data [Nogueras-Iso et al. 2021]. Metadata is efficiently utilized when organized into Metadata Standards [Hayslett 2023]. Standards act as frameworks that enable data interoperability [NISO 2017], such as the Data Catalog Vocabulary (DCAT) metadata standard, used as a framework to describe metadata in various domains [Nogueras-Iso et al. 2021]. Metadata standards can be classified as Descriptive, Structural, and Administrative [Formenton et al. 2018].

The literature presents various studies associated with metadata standards. In [Tauil 2018], a Systematic Literature Review (SLR) is described. The study aimed to map studies on digital preservation metadata in Cloud Services (CS). The specific objectives of the study included characterizing the functionalities and applications of CSs, examining the metadata standards applied in this context, and analyzing research trends.

In [Cerrao 2019], a study on conceptual models and description schemes applied in the bibliographic domain, specifically in Institutional Digital Repositories (IDRs), is presented. Using an SLR approach, the study established specific objectives related to IDR investigation, following an evaluation of the Functional Requirements for Bibliographic Records (FRBR) conceptual model and the guidelines contained in the Resource Description and Access (RDA). Among the results is the identification of theoretical and methodological elements for the application of FRBF and RDA in IDRs. Another SLR is presented in [James 2020]. The study identified metadata elements and metadata standards employed in the representation of government information. The author emphasizes that these standards and metadata elements could be employed for various purposes.

In the study conducted by [da Silva 2020], the documentation of artistic collections in national and international contexts was addressed. Among the recorded observations is that, on the international scene, there are more initiatives related to the topic compared to Brazil. The following international guidelines were analyzed and compared: Categories for the Description of Works of Art (CDWA), International Committee for Documentation of the International Council of Museums (CIDOC ICOM), and Standard Procedures for Collections Recording Used in Museums (SPECTRUM). The aim of the analysis and comparison was to select elements considered fundamental to propose a metadata schema that could serve as a basis for museological procedures.

3. Research Methodology

The methodology adopted for this study consists of the following stages: defining the research questions, developing the search strategy and selecting research sources, constructing the search string, and defining the selection criteria. Each stage is described below.

3.1. Research Questions

After defining the study's objective, the following research questions were established:

Table 1. Search strings used in the bibliographic databases.

Database	Query
ACM	“computer science” AND “metadata” OR “metadata standards”
Scopus	“computer science” AND (“metadata” OR “metadata standards”)
ScienceDirect	“computer science” AND “metadata” OR “metadata standards”
SpringerLink	“computer science” AND (“metadata” OR “metadata standards”)
IEEE	“computer science” AND “metadata” OR “metadata standards”

- **RQ1**: What metadata standards have been used in Computer Science?
- **RQ2**: Which research fields in Computer Science have considered the use of metadata standards?

3.2. Search Strategy and Research Sources

To address the research questions, a search strategy based on automated search was defined. The following databases were considered as research sources: IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, and Scopus. These databases are globally recognized in Computer Science, allowing for advanced queries and featuring studies with high-impact factors (JCR and/or H-index).

3.3. Search String

In this study, the search string used incorporated the following expressions or terms: ”computer science,” ”metadata,” and ”metadata standards.” These expressions were combined using logical operators ”AND” and ”OR.” Table 1 provides details of the search string applied across the databases. Table 1 presents the search strings used in the ACM, Scopus, ScienceDirect, SpringerLink, and IEEE databases. Variations in the composition of the strings can be observed due to the characteristics of each database.

3.4. Selection Criteria

During the reading, analysis, and selection process, different selection criteria were considered. These criteria assisted in the inclusion and exclusion of studies. Below, the inclusion and exclusion criteria are presented.

For **inclusion**, the following criteria were considered:

- Explicit citation of one or more Metadata Standards being used in Computer Science.
- Explicit citation of one or more research fields in Computer Science.

For **exclusion**, the following criteria were considered:

- Studies not within the context of Computer Science.
- No explicit citation of one or more Metadata Standards being used in Computer Science.
- No explicit citation of one or more research fields in Computer Science.
- Studies not described in English (impedes dissemination and reproducibility).
- Studies not published in conferences or journals.
- Duplicate studies.
- Studies unavailable, even after contacting the authors.
- Secondary studies.

4. Results

The study selection process considered activities such as reading the title and abstract, checking keywords, and applying selection criteria. At the end of the process, a set of selected studies was established. The selected studies are presented in Table 2.

Table 2. Selected studies

Title	Mentioned Standard
A generic framework for the integration of heterogeneous metadata standards into a multimedia information retrieval system	Dublin Core (DC), Large-Scale Concept Ontology for Multimedia (LSCOM), Digital Imaging and Communications in Medicine (DICOM), TV-Anytime Set, Exchangeable Image File Format (Exif), Extensible Metadata Platform (XMP), International Press Telecommunications Council Information Interchange Model (IPTC IIM), ISO/IEC 15938 - MPEG-7, MXF (Material Exchange Format), Text Encoding Initiative (TEI) e ID3 Standard
Metadata standards for data warehousing open information model vs. common warehouse metadata	Open Information Model (OIM), Common Warehouse Metamodel (CWM), Structured Query Language (SQL) Standard, Structured Query Language for Java Applications (SQLJ) Standard
Metadata Standards for Semantic Interoperability in Electronic Government	Dublin Core (DC)
The METAFOR project - preserving data through metadata standards for climate models and simulations	XML Metadata Interchange (XMI)
A Study on Various Semantic Metadata Standards to Improve Data Usability	Learning Object Model (LOM), Dublin Core (DC), Darwin Core (DwC), Linked Data (LD), Directory Interchange Format (DIF), Gateway of educational materials (GEM), Metadata object description schema (MODS), Friend of a Friend (FOAF) Standard, Web Ontology Language (OWL)
An Automatic Learning Contents Selector Based on Metadata Standards	IMS Global Leaming Consortium leaming object, Sharable Content Object Reference Model (SCORM), Learner Information Package (LIP), Simple Ontology Extension (SHOE), QTI (Question & Test Interoperability)
Customizing Discipline-based Metadata Standards for Digital Preservation of Living Epic Traditions in China	Dublin Core (DC)
Motion Imagery Metadata Standards Assist in Object and Activity Classification	Key-Length-Value (KLV), Motion Imagery Standards Board (MISB), STANAG 4609
Practical problems of implementing metadata standards in official statistics	ISO standards
Research on Information Metadata Standards of Knowledge Organization - A Case Study of Chinese Digital Library	Dublin Core (DC), Web Ontology Language (OWL)
Sar Metadata Standards - Single-Look Complex Data	Sensor Independent Complex Data (SICD), ISO 19115-1, ISO 19115-2, ISO 19130-1, ISO 19130-2, Synthetic Aperture Radar (SAR)
The role of metadata standards in EOSDIS data search and retrieval	NASA's Earth Observing System Data and Information System (EOSDIS), Directory Interchange Format (DIF), Hierarchical Data Format - Earth Observing System (HDF-EOS)
An international analysis of the extensions to the IEEE LOMv1.0 metadata standard	Learning Object Model (LOM)
Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS)	Medical Subject Headings (MeSH), Gene Ontology (GO)
Metadata standard interoperability - application in the geographic information domain	Dublin Core (DC), Machine Readable Cataloging (MARC), ISO 19115
Specifying metadata standards for metadata tool configuration	Dublin Core (DC), Government Information Locator Services (GILS)
EVENTSKG: A Knowledge Graph Representation for Top-Prestigious Computer Science Events Metadata	Resource Description Framework (RDF)
Analyzing the evolution of computer science events leveraging a scholarly knowledge graph: a scientometrics study of top-ranked events in the past decade	Resource Description Framework (RDF), and Web Ontology Language (OWL)
Computer Science: New Problems and Promising Methods (Review)	IEEE.802.15.4 Standard
The Role of Metadata Standards in EOSDIS Data Search and Retrieval	NASA's Earth Observing System Data and Information System (EOSDIS)

Title	Mentioned Standard
Metadata Feedback and Utilization for Data Deduplication Across WAN	Advanced Video Coding (AVC) or H.264
Improving Metadata Caching Efficiency for Data Deduplication via In-RAM Metadata Utilization	Advanced Video Coding (AVC) or H.264

Table 2 describes the selected studies in terms of their titles and the cited standards. The following sections will describe additional results associated with the established research questions.

4.1. Metadata Standards Used in Computer Science

Once the studies were selected, the stage of recording the necessary information for the research questions began. Considering the results presented in Table 2, specific characteristics can be observed. In addition to the application of standards, it was noted that different studies adopted models, schemas, metamodels, linked data, ontologies, languages, and packages for metadata standardization. Given this scenario, these results were considered, even if they do not exactly represent a metadata standard. Table 3 provides more information about the standards and structures used.

Table 3. Standards and structures used as metadata standards.

Standard	Description
Dublin Core (DC)	Set of metadata elements planned to facilitate the description of electronic resources
Extensible Metadata Platform (XMP)	Technology that allows embedding metadata into files during content creation
International Press Telecommunications Council Information Interchange Model (IPTC IIM)	Defines each field purposes and how values are stored as binary data
Text Encoding Initiative (TEI)	Consortium developing and maintaining a standard for digital text representation
Open Information Model (OIM)	Provides a clean and simplified model for XBRL, not supporting all XML-based syntax possibilities
Metadata object description schema (MODS)	MODS is a schema for a set of bibliographic elements that can be used for various purposes, especially for applications in libraries
Common Warehouse Metamodel (CWM)	Standard for exchanging metadata between modeling tools and databases
Structured Query Language for Java Applications (SQLJ) Standard	It consists of a translator and a runtime component and seamlessly integrates into your development environment
XML Metadata Interchange (XMI)	Integration framework for the exchange of models and any type of XML data. XMI is used in tool integration, repositories, applications, and data warehouses
Web Ontology Language (OWL)	Designed to represent rich and complex knowledge about things, groups of things, and relationships between things
IMS Global Learning Consortium Learning Object	Facilitate the discovery and retrieval of learning objects stored in more than one collection
Sharable Content Object Reference Model (SCORM)	Set of technical standards for eLearning software products
Learner Information Package (LIP)	Describes how students can interact with an online learning environment based on their preferences and needs
Simple Ontology Extension (SHOE)	Allows web page authors to annotate their web documents with machine-readable knowledge
Question and Test Interoperability (QTI)	Enables the exchange of item content and test and results data between authoring tools, item banks, test construction tools, learning platforms, assessment delivery systems, and scoring/analysis engines
Key-Length-Value (KLV)	KLV is a data encoding standard used to embed information or data in video feeds
Earth Observing System Data and Information System (EOSDIS)	Provides end-to-end capabilities for managing NASA Earth science data from various sources - satellites, aircraft, field measurements, and various other programs
Synthetic Aperture Radar (SAR)	Provides the means to describe the instrument, acquisition mode and operation, as well as the processing algorithms applied to the data and any parameters related to these algorithms
IEEE Learning Object Metadata (LOM)	Specifies a conceptual data schema that defines the structure of a metadata instance for a learning object
Resource Description Framework (RDF)	Facilitates data merging, even when the underlying schemas are different, and specifically supports schema evolution over time

Standard	Description
IEEE.802.15.4 Standard	Specifications for the physical layer and medium access control sublayer for low-rate wireless data connectivity with fixed, portable, and mobile battery-powered or very limited battery devices
Advanced Video Coding (AVC) or H.264	Metadata standard for video compression that provides video quality at low bit rates

The results presented in Table 3 allow for the analysis of standards, models, schemas, metamodels, and packages in terms of their contribution to metadata standardization.

4.2. Research Fields in Computer Science that Consider Metadata Standards

To answer RQ2, we conducted the association of studies that explicitly mentioned Metadata Standards in CS. The classification system adopted by ACM was considered (ACM, 2023). Table 4 presents the outcome of this association.

Table 4. Associations between studies and research fields in CS.

Subject	Article Title
Information Systems	"Metadata Standards for Semantic Interoperability in Electronic Government"
Computing Methodologies	"A generic framework for the integration of heterogeneous metadata standards into a multimedia information retrieval system"
Applied Computing	"An International Analysis of the Extensions to the IEEE LOMv1.0 Metadata Standard"
Networks	"A Deep Learning Classifier for Sentence Classification in Biomedical and Computer Science Abstracts"

In the association presented in Table 4, eight different research fields were considered. The research fields included Information Systems, Applied Computing, Computing Methodologies, and Networks. The same association was performed for subfields of research in CS. Table 5 presents the results.

Table 5. Associations between studies and subfields of research in CS.

Subfield	Article Title
Information Storage Systems	"Metadata Standards for Semantic Interoperability in Electronic Government"
Computer Graphics	"A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System"
Information Retrieval	"A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System"
Information Systems Applications	"A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System"
Computers in other Domains	"A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System"
Multimedia Information Systems	"A Generic Framework for the Integration of Heterogeneous Metadata Standards into a Multimedia Information Retrieval System"
Data Management Systems	"Metadata standards for data warehousing open information model vs. common warehouse metadata"
Education	"A Study on Various Semantic Metadata Standards to Improve Data Usability"
World Wide Web	"A Study on Various Semantic Metadata Standards to Improve Data Usability"
Life and Medical Sciences	"Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS)"
Network Protocols	"Computer Science: New Problems and Promising Methods (Review)"

In terms of subfields, various possibilities were observed, associated with Information Storage Systems, Computer Graphics, Information Retrieval, Information Systems Applications, Computers in other Domains, Multimedia Information Systems, Data Management Systems, Education, World Wide Web, Life and Medical Sciences, and Network

Protocols. In this association, it was noted that the same work was associated with different fields. This multiple association occurred because, in the respective study, different standards, models, and packages were presented.

5. Discussion of Results

From the presented results, it is possible to extract important information regarding the use of Metadata Standards in CC. The concept of standards considered by the selected studies can be interpreted as flexible, given that models, schemas, linked data, ontologies, languages, and packages were used to standardize metadata representation. This diversity of artifacts used for standardization purposes may indicate the absence of consolidated Metadata Standards in CC. Different standards were identified for different fields and subfields of research in CC. The classification adopted by ACM aided in associating studies with specific fields and/or subfields.

6. Threats to Validity

During the conduct of secondary studies, specific threats can be observed [Ampatzoglou et al. 2019]. In this study, the following categories of threats were considered and mitigated.

- Selection Validity: This category considers threats that could compromise the search and evaluation of studies. Examples of threats include the selection of databases and the construction of search strings. To minimize such threats, only databases deemed relevant to the fields of CS were considered [Nakagawa et al. 2017], and multiple tests were conducted with the established search string.
- Data Validity: This category considers threats that may be identified in data extraction and analysis. Examples of threats include errors that may occur during data collection and publication. To mitigate such threats, there was a separate and individual assessment by each of the authors of this study regarding the collected data.
- Research Validity: This category encompasses threats related to the overall research design. Examples of threats include the generalizability of results and the coverage of research questions. To minimize such situations, protocols established in the literature for Systematic Mappings were considered [Kitchenham et al. 2007]; [Petersen et al. 2008].

7. Final Remarks

This study presented a Systematic Mapping conducted to identify Metadata Standards for use in Computer Science. Initially, 43 studies were evaluated, of which 23 studies were selected for meeting the inclusion criteria. In addition to the explicit citation of standards, it was observed that other artifacts, such as models, schemas, linked data, ontologies, languages, and packages, were used to standardize the representation of metadata.

The authors understand that the results of the Systematic Mapping may indicate that standards and artifacts are considered in the description and/or representation of metadata, even if there is not yet a consolidated standard. Initial studies are being conducted

to investigate the application of Metadata Standards in specific fields, such as Software Engineering and Information Systems.

This project can be useful for studies related to metadata registration. It also can be relevant for studies that evaluate the utility and feasibility of metadata standards in specific applications.

Acknowledgements

Edson OliveiraJr thanks CNPq/Brazil Grant #311503/2022-5.

References

Acm, computing classification system. https://dl.acm.org/ccs.

Earth observing system data and information system (eosdis). https://www.earthdata.nasa.gov/eosdis.

Iim. https://iptc.org/standards/iim/.

Ims global learning consortium. learning object discovery and exchange. http://www.imsglobal.org/lode/index.html.

Ims question & test interoperability specification overview — ims global learning consortium. https://www.imsglobal.org/question/index.html.

Introduction to sar—arcgis pro — documentation. https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/introduction-to-synthetic-aperture-radar.htm.

Klvlib: Klvlib documentation. https://www.impleotv.com/content/klvlib/help/index.html.

Metadata object description schema: Mods (library of congress standards). https://www.loc.gov/standards/mods/.

Open information model. https://specifications.xbrl.org/spec-group-index-open-information-model.html.

Owl - semantic web standards. https://www.w3.org/OWL/.

Rdf - semantic web standards. https://www.w3.org/RDF/.

Resource description framework (rdf): Concepts and abstract syntax. https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

A review of metadata: a survey of current resource description formats - gils. http://www.ukoln.ac.uk/metadata/desire/overview/rev_10.htm.

Scorm explained 101: One minute scorm overview. https://scorm.com/scorm-explained/one-minute-scorm-overview/?utm_source=google&utm_medium=natural_search.

Shoe. https://www.cs.umd.edu/projects/plus/SHOE/.

Sqlj developer's guide. https://docs.oracle.com/en/database/oracle/oracle-database/23/jsqlj/index.html#Oracle.

Tei: Text encoding initiative. https://tei-c.org.

Al-Khalifa, H. S. and Davis, H. C. (2006). The evolution of metadata from standards to semantics in e-learning applications.

Ampatzoglou, A. et al. (2019). Identifying, categorizing and mitigating threats to validity in software engineering secondary studies. *Information and Software Technology*, 106:201–230.

Cerrao, N. G. (2019). *Análise das aplicações de metadados baseada em FRBR e RDA em repositórios institucionais digitais: contribuições do domínio bibliográfico*. PhD thesis, Universidade Federal de São Carlos.

da Silva, C. A. (2020). Esquema de metadados para descrição de obras de arte em museus brasileiros: uma proposta. https://www.teses.usp.br/teses/disponiveis/27/27151/tde-01032021-162722/pt-br.php.

Development, N. and Office, M. S. Marc standards (network development and marc standards office, library of congress). https://www.loc.gov/marc/.

Formenton, D. et al. (2018). Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. *Biblios: Journal of Librarianship and Information Science*, (68):82–95.

Gao, Y. et al. (2019). An efficient ring-based metadata management policy for large-scale distributed file systems. *IEEE Transactions on Parallel and Distributed Systems*, 30(9):1962–1974.

Hayslett, M. (2023). Libguides: Metadata for data management: A tutorial: Intro. https://guides.lib.unc.edu/metadata.

Izabel, M., Vendrusculo, L. G., and Melo, G. C. (2000). Metadados para a descrição de recursos de informação eletrônica: utilização do padrão dublin core. *Ciência da Informação*, 29(1):93–102.

James, T. D. (2020). Elementos de metadados para a descrição de jurisprudência: uma proposta aplicada em ambientes digitais jurídicos. *repositorio.ufscar.br*.

Kaur, A. et al. (2023). Literature review on metadata governance. *Open International Journal of Informatics*, 11(1):114–120.

Kitchenham, B. et al. (2007). Guidelines for performing systematic literature reviews in software engineering.

Lorihollasch. Mpeg-4 avc (h.264) - windows drivers. https://learn.microsoft.com/pt-br/windows-hardware/drivers/display/mpeg-4-avc--h-264-.

Marone, C. Synthetic aperture radar metadata content working group (p4002). https://www.grss-ieee.org/technical-committees/standards-for-earth-observations/working-group-standards-for-earth-observations/synthetic-aperture-radar-metadata-content-working-group-p4002/.

Mohan, A. B. K. and Choudhury, T. Introduction to sqlj. https://docs.oracle.com/en/database/oracle/

oracle-database/19/jsqlj/intro-to-SQLJ.html#GUID-7EDDF612-2C7C-4023-90F8-DCF291811D1B.

Nakagawa, E. Y. et al. (2017). Revisão sistemática da literatura em engenharia de software: teoria e prática.

NISO (2017). Understanding metadata: What is metadata, and what is it for?: A primer — niso website. https://www.niso.org/publications/understanding-metadata-2017.

Nogueras-Iso, J. et al. (2021). Quality of metadata in open data portals. *IEEE Access*, 9:60364–60382.

Petersen, K. et al. (2008). Systematic mapping studies in software engineering. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12*, pages 1–10.

Tauil, J. C. S. (2018). Metadados de preservação em cloud services. In *repositorio.ufscar.br*.

Weiss, M. (2009). Xml metadata interchange. In Liu, L. and Özsu, M. T., editors, *Encyclopedia of Database Systems*. Springer, Boston, MA.