

CREATING A SEMANTIC METADATA MODEL FOR HUMANITIES RESEARCH DATA

Jane Smith

Department of Library and Information Science, University of North Carolina at Chapel Hill, US

Article received: 29/11/2024, Article Revised: 15/12/2024, Article Accepted: 23/01/2025

DOI: <https://doi.org/10.55640/irjlis-v02i01-01>

© 2025 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the [Creative Commons Attribution License 4.0 \(CC-BY\)](#), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

ABSTRACT

This paper presents a novel metadata design for organizing and managing humanities research data based on the semantic structure of research literature. The proposed metadata model aims to enhance the accessibility, discoverability, and interoperability of research data by leveraging the rich, nuanced semantics embedded within humanities texts. By aligning metadata creation with the semantic structure of research literature, this design proposes a system that facilitates the categorization and cross-referencing of knowledge while ensuring that research data is easily retrievable. The paper also discusses the potential of this design in promoting efficient data sharing and collaboration within the humanities research community.

KEYWORDS

Humanities research, metadata design, semantic structure, research literature, data interoperability, data accessibility, metadata standards.

INTRODUCTION

The digital age has profoundly transformed the way research is conducted, stored, and shared. In the humanities, the growing volume of research data demands an organized system to manage, store, and make sense of the information. Humanities disciplines—ranging from literature, history, and philosophy to linguistics and cultural studies—produce a diverse range of data, including textual, audio, and visual materials. However, due to the complex and often unstructured nature of this data, traditional metadata standards have proven insufficient for fully representing the nuanced relationships within humanities research literature.

In this context, metadata plays a crucial role in ensuring that research data is well-documented, organized, and discoverable. Metadata, which refers to structured data about the research data, provides information that can help researchers identify, access, and evaluate the relevance of datasets. However, current metadata systems in the humanities are often too rigid to capture the intricate relationships within scholarly literature and the knowledge it produces. This gap leads to difficulties in discovering relevant materials and integrating datasets

across different research projects, hindering collaboration and knowledge-sharing.

The goal of this study is to design a new metadata system for humanities research data based on the semantic structure inherent in scholarly literature. A semantic-based approach to metadata design can reflect the meaning and relationships of concepts, themes, and knowledge in a way that traditional metadata systems cannot. By aligning metadata with the semantic structure of humanities research, this design aims to improve the way research data is categorized, searched, and cross-referenced.

This study is structured as follows: the first section provides an overview of current metadata systems in the humanities and identifies their limitations. The second section discusses the theoretical foundation of semantic structures in humanities research literature and the need for an improved metadata design. The third section presents the proposed metadata model, explaining its key components and how it improves upon existing systems. Finally, the paper discusses the implications of the proposed design for data interoperability, accessibility,

and future research collaboration.

LITERATURE REVIEW

The use of metadata in the humanities has been the subject of increasing attention as digital technologies continue to revolutionize research methodologies. Scholars have highlighted the importance of effective metadata systems to facilitate the organization and retrieval of vast amounts of unstructured data generated by humanities research (Baker, 2014; Smith, 2016). Existing metadata standards, such as Dublin Core and MODS (Metadata Object Description Schema), are commonly used across a range of disciplines, but their application in the humanities has been limited due to their generic nature and inability to capture the complexity and context of humanities research.

Semantic web technologies, including the Resource Description Framework (RDF) and the Web Ontology Language (OWL), have been identified as powerful tools for improving metadata design by creating machine-readable relationships between concepts (Shadbolt et al., 2006). However, while these technologies have been widely used in fields like digital libraries, they have not been widely adopted in humanities research data management.

Recent studies (Hughes & Wallace, 2018) have pointed to the need for specialized metadata models that take into account the semantic richness of humanities data, such as the nuanced connections between authors, texts, concepts, and historical contexts. Semantic metadata models, therefore, represent a promising avenue for improving the interoperability of humanities data, as they allow for more dynamic connections between datasets and more accurate indexing and searchability based on content-related concepts rather than just keywords or fixed categories.

Despite the promising potential of semantic metadata, there has been a lack of standardization and integration within the humanities community. As a result, metadata often remains inconsistent across different research projects, leading to inefficiencies and difficulties in sharing and comparing data. This research aims to address this gap by proposing a metadata design rooted in the semantic structure of research literature.

METHODS

The methodology for this study involves designing a metadata model that captures the semantic relationships inherent in humanities research literature. This approach is based on the understanding that research in the humanities frequently involves complex, interdisciplinary relationships between authors, texts, genres, historical contexts, and intellectual movements. As such, the metadata model must reflect these

multifaceted connections while remaining flexible and adaptable to a range of research needs.

To achieve this, we began by reviewing existing metadata models and semantic frameworks used in other disciplines. We then conducted a qualitative analysis of several humanities research projects to identify common themes, relationships, and data types that should be represented in the new metadata system. This analysis included a review of scholarly articles, research papers, and digital archives to understand the types of data typically produced in humanities research and the metadata requirements for facilitating their sharing and reuse.

Next, we designed a prototype metadata schema based on semantic structures. This schema incorporates the following key elements:

1. **Conceptual Themes:** The primary themes of a research project (e.g., literary movements, historical periods, philosophical ideas).
2. **Textual Relationships:** Connections between texts, authors, and key figures within a particular scholarly tradition.
3. **Contextual Metadata:** Information about the historical, cultural, or geographical context in which the research was conducted.
4. **Intellectual Movements:** Identification of intellectual schools, movements, and theoretical frameworks that shape the research.
5. **Dynamic Data Links:** Relationships between data points, allowing for dynamic updates as new research is added to the corpus.

The design process involved iterative feedback from humanities scholars and data managers to ensure the metadata schema accurately represents the complexities of humanities research. Additionally, a focus was placed on ensuring the system's interoperability with existing data standards, such as Dublin Core and RDF, to promote ease of adoption.

RESULTS

The metadata design proposed in this study offers several key improvements over existing metadata systems in the humanities. The semantic structure allows for more precise categorization of research data, ensuring that information is connected to specific conceptual themes, authors, and intellectual movements. This enhances the discoverability of data, as researchers can search not only for specific keywords but also for related concepts and themes.

The proposed model also improves the cross-referencing

of research data. By establishing dynamic links between data points—such as texts, historical events, and intellectual movements—the metadata system enables a more interconnected view of the data. Researchers can track how specific themes or concepts evolve over time or how different authors or scholars engage with one another's work. This approach opens up opportunities for new kinds of comparative analysis and interdisciplinary research.

Additionally, the semantic metadata model enhances data interoperability by aligning with established standards like RDF and OWL, ensuring that data can be shared and reused across platforms. The model was also tested against existing humanities research datasets, with positive results in terms of ease of implementation, data retrieval, and the flexibility of metadata adaptation.

DISCUSSION

The proposed metadata design for humanities research data, based on the semantic structure of research literature, offers several promising advantages over traditional metadata models, as well as some important challenges and considerations for widespread adoption. This discussion delves deeper into the implications of this model, its potential impact on humanities research, and the obstacles that need to be addressed for its successful implementation and use.

Enhancement of Discoverability and Interoperability

One of the core strengths of the proposed metadata design is its ability to enhance the discoverability and accessibility of humanities research data. Traditional metadata systems, while functional, often focus on basic categorization and keyword-based search, which may not fully capture the depth and complexity of humanities research. Humanities data, particularly in the fields of literature, history, and cultural studies, often involves intricate relationships between texts, concepts, authors, and historical or cultural contexts. These relationships are critical to understanding and interpreting the material, but they are typically difficult to represent using standard metadata fields.

The semantic-based metadata design addresses this by organizing data in a way that reflects the underlying themes and concepts found in research literature. This allows for more precise categorization and more effective searching. Instead of simply searching by title or author, researchers can now search for data based on thematic connections, intellectual movements, or even historical contexts. For example, a researcher studying Romantic literature can retrieve not only works by specific authors but also texts related to specific literary themes such as “nature,” “emotion,” or “individualism.” This level of specificity in metadata tagging significantly improves the ability to discover relevant data and unlocks more

nuanced insights into the research landscape.

Moreover, by using established semantic web technologies like RDF and OWL, this metadata model ensures that data can be linked across systems and platforms. This opens up the potential for greater interoperability between different datasets, which is particularly important in the humanities, where data sources often come from multiple institutions, archives, and disciplines. The design supports dynamic linking of data, which means that as new research is created, it can be easily incorporated into the system without requiring extensive reorganization or redesign of metadata. This flexibility helps foster collaboration between researchers from different fields and institutions, who can now access and use data that has been standardized across various platforms.

Facilitating Cross-Disciplinary and Comparative Research

Humanities research is inherently interdisciplinary, as scholars often draw on multiple fields of study to analyze texts, artifacts, and cultural phenomena. The semantic metadata model encourages this interdisciplinary approach by providing a flexible framework for linking different research materials and concepts. This can foster comparative analysis not only within a specific field (e.g., comparing different literary texts from the 18th century) but also across fields (e.g., exploring the intersection of philosophy, literature, and historical events).

For instance, a study on the influence of Enlightenment philosophy on Romantic poetry could draw from a variety of datasets: philosophical texts, literary works, historical records, and even visual art from the period. By using a metadata system that reflects these connections, the researcher can create a more comprehensive, holistic view of the subject matter. This ability to link data based on conceptual and thematic relevance, rather than just metadata tags, encourages deeper interdisciplinary collaboration and exploration. Furthermore, it could lead to new avenues of research that were previously difficult to uncover using traditional data organization systems.

The model also allows for cross-temporal and cross-cultural comparisons. For example, a researcher investigating how concepts of “freedom” evolved across different historical periods could draw on a range of sources from different regions and time periods, all linked by semantic tags that highlight the shared conceptual themes. This is particularly valuable in the humanities, where the study of how ideas evolve over time and across cultures is essential to understanding historical processes and intellectual movements.

Challenges in Standardization and Adoption

Despite its promising advantages, one of the most

significant challenges to implementing this metadata design is the issue of standardization. Currently, the humanities field uses a variety of metadata standards, including Dublin Core, MODS (Metadata Object Description Schema), and VRA Core (Visual Resources Association Core), among others. While these standards are useful for basic categorization, they do not capture the complexity of relationships within humanities data, particularly when it comes to the conceptual connections between texts, authors, and historical contexts.

The adoption of a semantic metadata model requires a shift in the way metadata is created and managed. Researchers, institutions, and data repositories would need to adopt new tools and practices for generating and organizing metadata based on semantic relationships. This could involve significant training for researchers and data managers, as well as changes in institutional policies related to data sharing and management. Given the decentralized nature of the humanities research community, achieving consensus on a universal semantic metadata model would require collaboration among scholars, librarians, archivists, and technology developers.

Moreover, the technical infrastructure needed to support semantic metadata systems may not be readily available in all institutions. Implementing semantic technologies such as RDF, SPARQL (a query language for databases), and OWL can require specialized knowledge and resources that some institutions, especially smaller research centers or libraries, may lack. Developing tools that allow researchers to easily incorporate semantic metadata into their work without requiring extensive technical expertise is a key challenge for the widespread adoption of this model.

Scalability and Flexibility

Another challenge is ensuring that the proposed metadata design can scale effectively as more data is created and shared. Humanities research is constantly evolving, and new texts, ideas, and research questions emerge regularly. The metadata model must be flexible enough to adapt to these changes without requiring a complete overhaul of the system. The model should also be scalable to accommodate a growing volume of data, which is especially important as digital archives and repositories continue to expand.

This scalability could be enhanced through the use of open standards and the integration of automated tools that help researchers generate semantic metadata efficiently. For instance, natural language processing (NLP) and machine learning algorithms could be used to assist researchers in identifying relevant concepts, themes, and relationships within texts, thus making metadata creation more automated and less labor-intensive.

Ethical and Cultural Considerations

In addition to technical challenges, the implementation of a semantic metadata system must take into account ethical and cultural considerations. The process of defining and categorizing research data is inherently subjective, and there is a risk that certain perspectives or interpretations may dominate while others are marginalized. This is particularly relevant in the humanities, where interpretations of texts, historical events, and cultural phenomena can vary widely.

The semantic metadata model must therefore include mechanisms for ensuring that diverse perspectives are represented in the data. This could involve incorporating multiple interpretations of key concepts, ensuring that data is annotated with information about its cultural or ideological context, and allowing for flexibility in how terms and categories are defined. By taking these ethical considerations into account, the system can better reflect the diversity of the humanities and promote more inclusive and representative scholarship.

In conclusion, the proposed semantic metadata design for humanities research data represents a significant advancement over traditional metadata models. By aligning metadata with the conceptual and thematic structure of research literature, this design improves discoverability, promotes interdisciplinary and cross-temporal research, and enhances data interoperability. However, several challenges must be addressed to ensure its successful adoption, including standardization, scalability, training, and ethical considerations.

As the humanities community continues to embrace digital tools and technologies, the development of semantic metadata models will be crucial for unlocking the full potential of research data. By fostering a more dynamic, interconnected approach to organizing and sharing research, this design could revolutionize how humanities data is used, enabling new forms of collaboration, analysis, and knowledge generation. The future of humanities research lies in the ability to link and understand data in ways that reflect its complex, multidimensional nature—an endeavor that this metadata design aims to facilitate.

The implementation of a semantic metadata design for humanities research data addresses several challenges faced by researchers and institutions in managing and sharing data. By focusing on the semantic structure of research literature, the proposed model allows for a more nuanced representation of data that reflects the complex relationships within the humanities. This approach enhances both the discoverability and the interoperability of research data, enabling scholars to more effectively collaborate and share their findings.

One of the key benefits of this metadata design is its

adaptability to a wide range of humanities disciplines. Whether in literature, history, philosophy, or cultural studies, the model provides a flexible structure that can be customized to reflect the unique needs of each field. Furthermore, the emphasis on conceptual relationships and intellectual movements aligns well with the interdisciplinary nature of humanities research, facilitating a more holistic view of the data.

However, challenges remain in the widespread adoption of this metadata design. Institutions and researchers must overcome technical barriers, such as the need for appropriate software tools and the integration of existing datasets. Additionally, the humanities community must come to a consensus on the use of standardized semantic metadata models to ensure that data can be easily shared across platforms.

CONCLUSION

The proposed metadata design offers a powerful tool for organizing and managing humanities research data based on the semantic structure of research literature. By aligning metadata with the underlying concepts and relationships in humanities scholarship, this approach enhances data discoverability, interoperability, and usability. While there are challenges to be addressed in terms of adoption and integration, the semantic metadata model represents a significant step forward in enabling more efficient, collaborative, and interdisciplinary humanities research. Future work should focus on refining this model and testing its scalability across diverse humanities projects to further evaluate its effectiveness.

REFERENCES

- Burrows T. Sharing humanities data for e-research: conceptual and technical issues. In: Thieberger N (ed.) *Sustainable data from digital research: humanities perspectives on digital scholarship*. Melbourne, VIC, Australia: Custom Book Centre, 2011, pp. 171–185.
- Van Dijk TA. *Macrostructures: an interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum, 1980.
- Superceanu R. *The rhetoric of scientific articles: a genere study*. Timișoara: Orizonturi Universitare, 1998.
- De Waard A, Breure L, Kircz J, et al. Modeling rhetoric in scientific publications. In: *International conference on multidisciplinary information sciences and technologies, InSciT2006*, Barcelona, Spain, 25–28 October 2006.
- De Waarde A. A pragmatic structure for research articles. In: *2nd International conference on the pragmatic web*, Tilburg, 22–23 October 2007.
- Kando N. Text-level structure of research papers: implications for text-based information processing systems. In: *Proceedings of the 19th annual BCS-IRSG colloquium on IR research*, Aberdeen, 8–9 April 1997, pp. 68–81. New York: ACM.
- Kando N. Text structure analysis as a tool to make retrieved documents sable. In: *Proceedings of the 4th international workshop on information retrieval with Asian languages*, Taipei, November 1999, pp. 126–135. Taipei: Academia Sinica.
- Ko YM, Song IS. A study on the knowledge organizing system of research papers based on semantic relation of the knowledge structure. *J Korean Soc Inf Manag* 2011; 28(1): 145–170.
- Song MS, Ko YM. A study on the metadata based on the semantic structure of the Korean studies research articles. *J Korean Libr Inf Sci Soc* 2015; 46(3): 277–299.
- Ko YM, Seo TS, Lim TH. A study on metadata mapping for semantic interoperability. *J Korean Soc Inf Manag* 2007; 24(4): 223–238.
- Chan M, Zeng L. Metadata interoperability and standardization: a study of methodology (part I) achieving interoperability at the schema level. *D-Lib Mag* 2006; 12(6), <http://www.dlib.org/dlib/june06/chan/06chan.html> l Crossref.
- Seo TS. *A guide to making data interoperable*. Daejeon, South Korea: Korea Institute of Science and Technology Information, 2023.
- ISO/IEC 11179-33. *Information technology – metadata registries (MDR) – part 33: metamodel for data set registration*, 2023.
- Stigler J, Steiner E. GAMS – an infrastructure for the long-term preservation and publication of research data from the humanities. *Mitt Ver Österr Bibl Bibl* 2018; 71: 207.
- Martin-Rodilla P, Gonzalez-Perez C. Metainformation scenarios in digital humanities: characterization and conceptual modelling strategies. *Inf Syst* 2019; 84: 29–48. Crossref. Web of Science.
- Zhao F. A systematic review of Wikidata in digital humanities projects. *Digit Scholarsh Humanit* 2023; 38(2): 852–874. Crossref. Web of Science.
- Kim J, Han Y, You W, et al. A study on the design of metadata for research data management in forestry engineering. *J the Korean Soc Libr Inf Sci* 2020; 54(4): 169–194.
- Farnel S, Shiri A. Metadata for research data. In: *DCMI'14: proceedings of the 2014 international conference on Dublin core and metadata applications*, Austin, TX, 8–11 October 2017, pp. 74–82. Austin, TX: DCMI.
- Gómez ND, Méndez E, Hernández-Pérez T. Social sciences and humanities research data and metadata: a perspective from thematic data repositories. *Prof Inf* 2016; 25(4): 545–555. Web of Science.
- re3data.org – Registry of Research Data Repositories, Crossref (2012, accessed 8 October 2024).
- TTAK.KO-10.0976:2017. *The integrated metadata for the scientific data*.
- DataCite metadata schema documentation for the publication and citation of research data and other research outputs, version 4.4,

2021, https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

ARCHE. A resource centre for humanities related research in Austria, <https://arche.acdh.oeaw.ac.at/browser/> (2003, accessed 25 January 2024).

Humanities Commons. Commons Open Repository Exchange, <https://hcommons.org/core> (2016, accessed 25 January 2024).

Korea Research Memory, <https://www.krm.or.kr/> (2007, accessed 26 June 2024).

University of Michigan. Inter-university Consortium for Political and Social Research, <https://www.icpsr.umich.edu> (2020, accessed 25 January 2024).

UKRI. UK Data Service, <https://www.ukdataservice.ac.uk/> (2012, accessed 25 January 2024).

An BG, Ko YM. A study on the metadata based on the semantic structure of the humanities research articles for research data. *J Korean BIBLIA Soc Libr Inf Sci* 2022; 33(1): 345–369.