ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Representation of SSH Research Data: Comparing Typological Structures in Zenodo and re3data

Isabella Henrique Lima Pereira

Fluminense Federal University, Brazil

Suellen Oliveira Milani

Fluminense Federal University, Brazil

Abstract

This study examines challenges in organizing and categorizing research data produced in the Social Sciences and Humanities. Due to their heterogeneous and context-dependent nature, Social Sciences and Humanities data can bring challenges for standardized representation in research data infrastructures. The study analyzes data typologies adopted by the Zenodo repository and the re3data registry to identify how these platforms categorize research outputs. The research adopts an exploratory approach combining literature review and analysis of the classification structures available in each platform. The analysis reveals differences in the typologies used and highlights limitations in representing the diversity of Social Sciences and Humanities research data. The findings highlight the need for flexible knowledge representation approaches that accommodate the epistemological and methodological diversity of SSH research practices

1.0 Introduction

The expansion of open science practices has intensified the production, sharing, dissemination, and preservation of research data, positioning data infrastructures as central components of contemporary scholarly communication. Repositories, directories, and other data services increasingly mediate how research outputs are organized, discovered, and reused, raising important questions about the representation of disciplinary knowledge within these systems.

Research data are not produced uniformly across disciplines. In the Social Sciences and Humanities (SSH), data frequently emerge from qualitative and context-dependent methodologies such as interviews, ethnographic observations, archival documents, textual corpora, and audiovisual materials. These characteristics make SSH data heterogeneous and strongly connected to sociocultural contexts, creating challenges for description, classification, and preservation within standardized infrastructures. Consequently, infrastructures designed for large-scale data sharing often rely on generalized typologies that may inadequately represent disciplinary diversity.

This challenge is intensified by the absence of consensus regarding what constitutes "data" in humanities research. Lin (2023) demonstrated that humanities researchers frequently attribute distinct meanings to data, research materials, sources, and documents, revealing conceptual ambiguities that complicate data curation and classification efforts. Because research data infrastructures rely on classification mechanisms to organize and provide access to information, questions concerning representation and categorization become central analytical concerns.

Knowledge Organization (KO) provides theoretical tools to examine how research data are classified and represented. Classification decisions reflect epistemological assumptions and disciplinary traditions rather than purely technical choices. Analyzing these classification structures reveals tensions between standardization and diverse research practices.

Despite the growing adoption of research data infrastructures, it remains unclear whether the classification typologies embedded in these systems adequately represent the epistemological and methodological diversity of SSH research data. Therefore, this study addresses the following research question: To what extent do the classification structures adopted by Zenodo and re3data adequately represent SSH research data?

The general goal of this research is to analyze the classification typologies adopted by Zenodo and re3data regarding their capacity to represent SSH research data. Specifically, the study aims to: (1) identify the typologies adopted by each platform; (2) compare their classification logics; and (3) discuss their implications for the representation of SSH research data from a Knowledge Organization perspective.

This investigation constitutes broader doctoral research regarding how SSH researchers manage and share research data, with future empirical steps focusing on researchers' practices and perceptions.

2.0 Research Data and Knowledge Organization

The growing importance of research data in scholarly communication has intensified discussions about their terminology and organization. Open science initiatives emphasize accessibility and reuse, positioning research data as key elements of knowledge production (OECD 2007). Research data are commonly described as materials necessary to validate research findings (OECD 2007), yet in SSH this definition becomes complex due to interpretive methodologies. Borgman (2015) argues that data are products of research contexts shaped by disciplinary norms. SSH data may include interviews, field notes, archival records, images, and audiovisual materials whose meaning depends on interpretation.

Recent studies have further demonstrated the diversity of research data in the Humanities. Gualandi et al. (2023), based on interviews with humanities researchers, identified thirteen distinct categories of research data, ranging from archival materials and textual *corpora* to publications and digital resources. Their findings reinforce the difficulty of establishing

uniform classifications capable of representing the diversity of SSH research practices in a defensible way.

Information practices differ among research communities and influence how data are produced and shared (Huvila, Isto, and Luanne Sinnamon 2022; Given and Rebekah 2018). SSH researchers often perceive data as inseparable from analytical processes, challenging standardized infrastructures.

From a KO perspective, classification systems represent intellectual structures reflecting theoretical assumptions about knowledge domains (Hjørland 2017). Research data infrastructures operate as knowledge organization systems that enable organization and retrieval through metadata and typologies (Zeng and Chan 2004). While standardization supports interoperability, it may insufficiently represent disciplinary diversity, particularly in SSH contexts.

Recent investigations of Humanities data repositories have highlighted both the diversity of humanities datasets and the persistent challenges of achieving FAIR representation. Poljak Bilić and Posavec (2024) emphasize that humanities research data encompass a broad range of materials and often resist standardized definitions, reinforcing the need for infrastructures capable of accommodating heterogeneous forms of knowledge production.

Examining classification structures in research infrastructures therefore allows identification of epistemological assumptions embedded in digital systems and evaluation of their adequacy for representing diverse research practices.

3.0 Methodological Approach

This study adopts an exploratory qualitative approach to examine how research data are represented within large-scale infrastructures, focusing on classification structures embedded in platforms rather than on user practices. The research employs documentary analysis, a method suitable for examining institutional systems and informational structures (Bowen 2009). Public interfaces and categorical and typological structures of two platforms were analyzed: Zenodo, a multidisciplinary repository, and re3data, an international registry of research data repositories.

In this paper, the term ‘classificatory structures’ is used analytically to refer to the sets of categories, typologies, and controlled terms through which platforms organize, describe, and make research outputs or repositories discoverable. These structures are not treated as formal classification systems in the strict sense, but as ‘organizational devices’ that perform classificatory functions within research data infrastructures.

The selection of Zenodo and re3data was based on their international relevance and complementary roles within the research data ecosystem. Zenodo was chosen because it is one of the largest multidisciplinary repositories supporting direct deposit and categorization of research outputs. Re3data was selected because it is a leading

international directory of research data repositories, providing structured descriptions of data infrastructures across disciplines

Although Zenodo and re3data operate at different levels of the research data ecosystem, the purpose of this comparison is not to establish equivalence between datasets and repositories. Rather, the analysis focuses on the classification structures embedded in both infrastructures as KO mechanisms. The comparison between these platforms enables the examination of two distinct classification layers: the representation of research outputs and the representation of repositories. This perspective allows examination of how different infrastructural layers represent research data and related informational entities.

This comparative design extends previous single-platform analyses of re3data by examining how classificatory structures operate across different infrastructural layers. Rather than focusing exclusively on the content-type categories of one directory, the present study compares the representational logics of a repository and a registry, thereby broadening the analytical scope of the investigation.

The analysis involved three stages: (a) identifying available typologies, (b) organizing categories for comparison, and (c) qualitatively interpreting similarities and limitations in the representation of SSH data. Rather than evaluating technical performance or FAIR compliance, this study understands these classificatory structures as socio-technical constructs shaped by epistemological and institutional factors (Hjørland 2017). This reflection is part of a broader doctoral thesis about SSH researchers' data management practices in Brazil, as mentioned above. Future steps include surveys and semi-structured interviews with researchers and analyses of discipline-specific repositories.

4.0 Data Typologies in Research Data Platforms

Research data infrastructures rely on classification structures to organize research outputs and enable discovery, access, and reuse. These structures often take the form of predefined typologies that guide users during data deposition or repository identification processes. In order to examine how research data are categorized within large-scale infrastructures, this study analyzed the data typologies presented in the public interfaces of the platforms Zenodo and re3data.

Zenodo is a multidisciplinary repository that allows researchers to deposit a wide range of research outputs, including datasets, publications, and other scholarly materials. During the submission process, users are required to assign a resource type to their upload, selecting from a predefined list of categories available in the platform interface, as demonstrated in Figure 1.

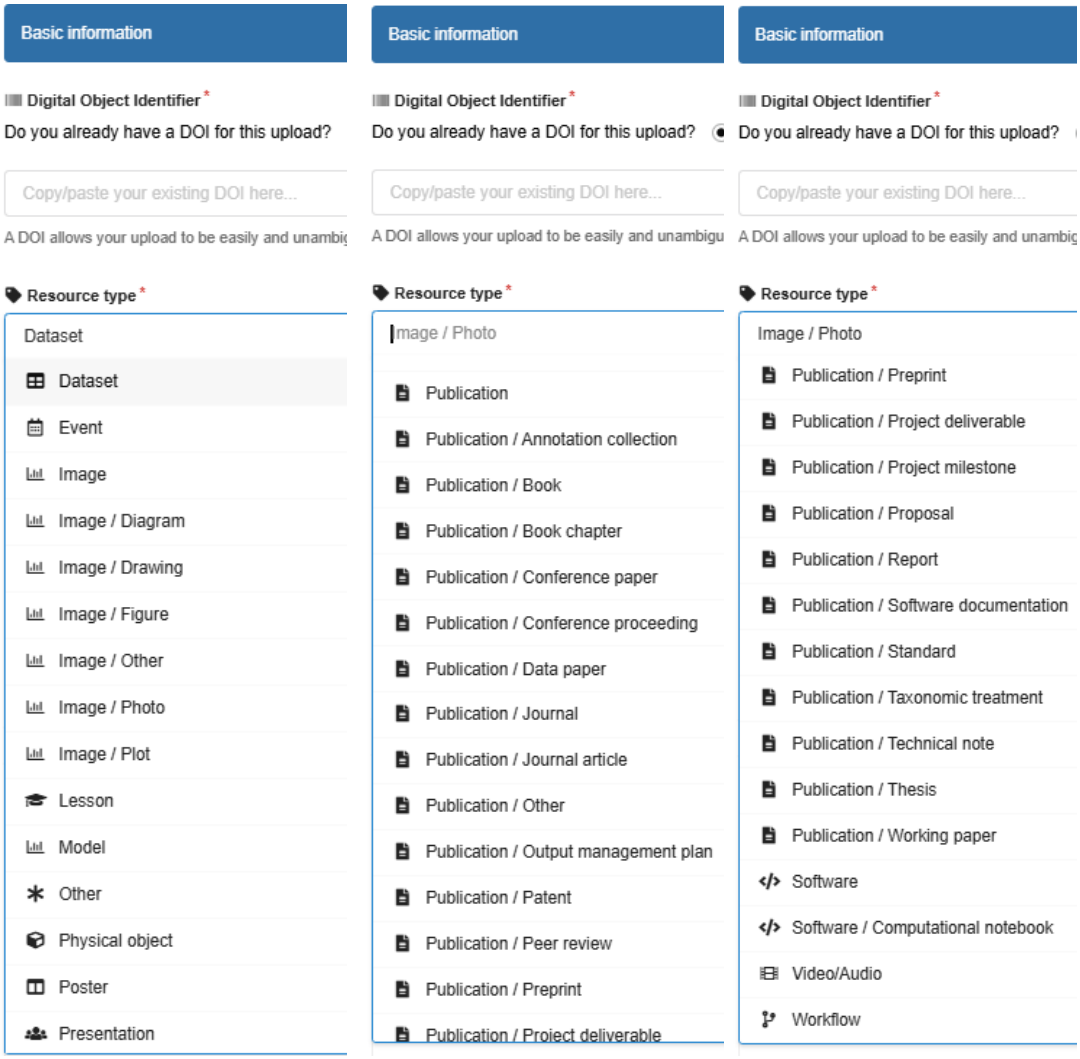

Figure 1. Resource-type categories used in Zenodo's upload process.
Source: Zenodo public interface.

The analysis identified that Zenodo organizes research outputs through a hierarchical typology that includes categories such as dataset, publication and other research outputs. Although these categories are designed to include multiple disciplines, they reflect a generalized approach to classification intended to support interoperability across scientific domains.

From a KO perspective, this typology prioritizes format-oriented classification rather than disciplinary or methodological specifications. Categories emphasize the material or technical nature of the resource rather than the epistemological characteristics of the research process. For example, qualitative interview data, ethnographic records, and statistical datasets may all be grouped under the same "dataset" category despite substantial differences in production context and analytical use. This suggests that interoperability is achieved at the expense of epistemological differentiation, privileging technical consistency over representational specificity.

Unlike Zenodo, re3data does not function as a repository for data deposition but as an international registry that indexes research data repositories. Its classification structure is therefore oriented toward describing repositories rather than individual datasets. Nevertheless, the platform employs structured categories that indirectly represent data typologies through disciplinary classifications and repository characteristics. The analysis of the public interface revealed that re3data organizes repositories primarily according to subject areas and data-related features, including disciplinary coverage, content types, and access conditions. Subject categories follow broad academic domains, enabling users to identify repositories associated with specific research fields, as illustrated in Figure 2.

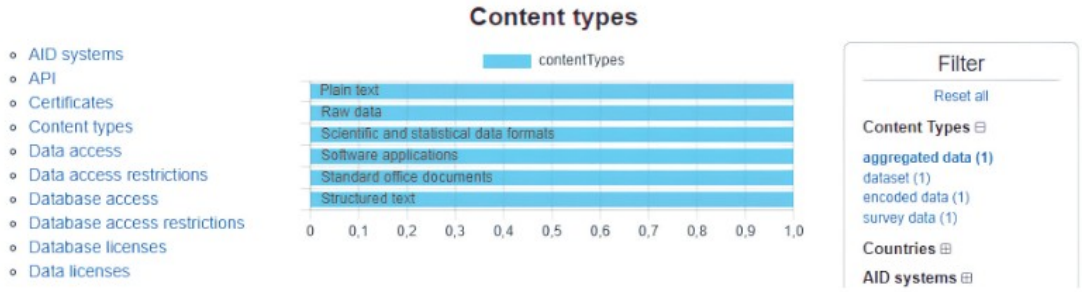

Figure 2. Example of content-type filtering structure in re3data.
Source: Reproduced from Pereira and Milani (2025, 471).

While this approach introduces disciplinary differentiation absent in Zenodo's resource typology, it still relies on standardized classifications that aggregate diverse SSH research practices into broad subject categories. Such classification may support navigation and interoperability but can reduce the visibility of methodological diversity within SSH domains. From a KO perspective, re3data operates at a meta-level of classification, where repositories, rather than datasets, constitute the primary units being organized. In this sense, the platform structures information about data infrastructures instead of individual research outputs. This distinction illustrates how classification operates across multiple layers of the research data ecosystem.

The comparison between Zenodo and re3data reveals complementary but distinct classification logics. Zenodo applies a resource-oriented typology focused on formats and output types, whereas re3data adopts a repository-oriented classification structured around disciplinary domains and infrastructural characteristics. Both approaches demonstrate the tension between standardization and defensible representation frequently discussed in KO research. Standardized categories enable interoperability and usability at scale, yet they may insufficiently capture the epistemological diversity of SSH research data. The absence of more specific categories related to qualitative methods, contextual materials, or interpretive data practices suggests limitations in how existing infrastructures represent SSH knowledge production.

These findings indicate that research data typologies embedded in digital platforms are not neutral organizational tools but reflect design decisions shaped by technical, institutional,

and epistemological considerations. Examining such classification structures provides insight into how KO principles are operationalized within research data infrastructures and how they may influence the visibility and accessibility of SSH research outputs.

5.0 Discussion: Challenges for Knowledge Organization

The comparison between Zenodo and re3data reveals a shared tendency toward classification generalization. Although both platforms support interoperability at scale, their structures prioritize standardization over epistemological specificity.

This tendency may also reflect broader assumptions associated with data-driven research cultures. Huvila and Sinnamon (2024) argue that not all research communities operate according to data-centered logics and identify alternative data-decentered and non-data cultures. Such perspectives are particularly relevant for SSH domains, where research practices are frequently grounded in interpretation, contextualization, and qualitative inquiry rather than standardized notions of data. Consequently, classification systems designed to support large-scale interoperability may inadequately capture forms of knowledge production that do not conform to data-driven assumptions.

Classification systems are socially situated constructions shaped by institutional and technical decisions (Bowker and Star 1999). In multidisciplinary infrastructures, broad schemas are adopted to accommodate heterogeneous outputs, often reducing domain specificity.

Zenodo illustrates this dynamic through format-oriented categories that treat research outputs as technically comparable objects. While enabling interoperability, such organization obscures differences in knowledge production practices. Distinct SSH datasets produced through interviews, ethnography, or archival interpretation become indistinguishable at the classification level. This limitation becomes particularly evident considering the diversity of SSH data types documented in recent studies.

Gualandi et al. (2023) identified thirteen distinct categories of research data in humanities research, illustrating the heterogeneity of materials, methods and outputs involved in knowledge production. Such diversity is only partially reflected in the broad categories adopted by large-scale infrastructures.

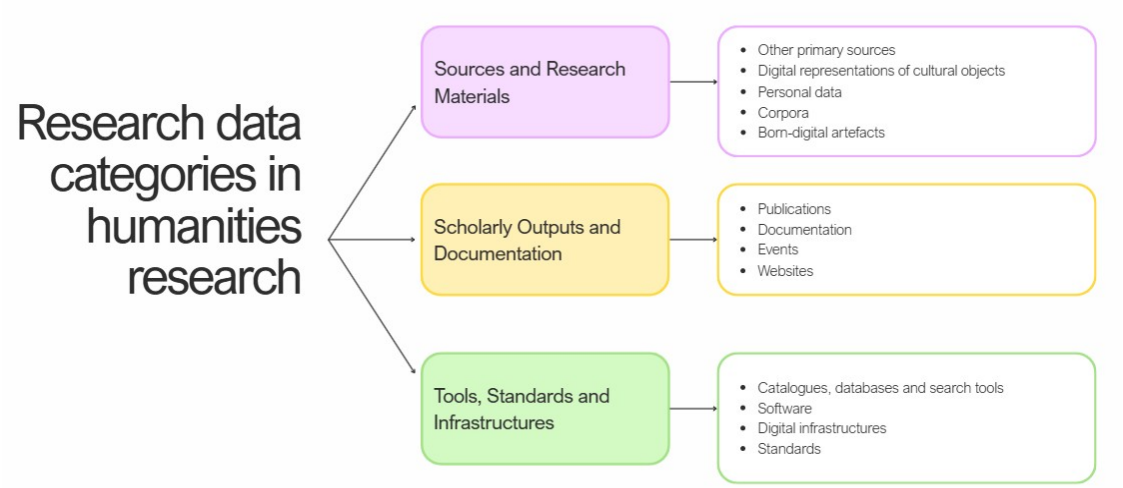

Figure 3. Heterogeneity of research data categories in humanities research. Source: Adapted from Gualandi et al. (2023).

Figure 3 shows that humanities research data extend across multiple dimensions of scholarly work, including research sources, communicative outputs, and digital or infrastructural elements. This heterogeneity helps explain why generalized categories adopted by large-scale infrastructures may be insufficient to represent SSH data with adequate domain sensitivity.

Hjørland (2017) emphasizes that domain-sensitive organization is essential for meaningful representation. When representation neglects epistemological contexts, discoverability and reuse may be limited, particularly in SSH research where interpretation and context are central.

Re3data introduces disciplinary representation but operates at a meta-level describing repositories rather than datasets. Although this reduces some generalization, epistemological diversity remains aggregated within broad domains.

These findings suggest that research data infrastructures actively shape knowledge visibility through the representation choices embedded in their organizational structures. This observation is consistent with previous work on re3data, which showed that its ‘content type’ categories adopt a generalist and universal approach to data typologies and that Knowledge Organization Systems have epistemological and cultural dimensions that shape how knowledge is produced and accessed (Pereira and Milani 2025).

From a Knowledge Organization perspective, representation choices are not merely a technical mechanism for retrieval but a representational practice that influences what becomes visible, searchable and reusable.

Rather than rejecting standardization, the results indicate the need for multi-layered representation approaches capable of balancing interoperability with domain sensitivity and epistemological diversity. Such approaches can be developed through the collaboration of information professionals, disciplinary experts, representatives of SSH

research communities, repository managers, and repository users, allowing categories and metadata structures to be informed by both Knowledge Organization principles and the actual research practices of SSH communities. Possible strategies include consulting librarians and data curators in the design of metadata schemas, controlled vocabularies, and descriptive guidelines; involving researchers from specific SSH fields in the validation of categories, examining terms commonly used in deposited datasets, and comparing these terms with categories already used to describe other forms of scholarly documents.

This dialogical dimension is particularly relevant because research data infrastructures are maintained by professional and institutional communities that may be open to clarification, feedback, and refinement. Exploratory contact with re3data staff during the research process, for example, indicated that the logic underlying platform categories can be discussed and clarified, suggesting possibilities for collaborative improvement rather than merely external critique. In this sense, improving representation would not require abandoning standardized infrastructures, but rather expanding them through iterative, community-informed, and context-sensitive mechanisms.

6.0 Conclusion

This study examined how research data are represented within two widely used platforms in the research data ecosystem, the repository Zenodo and the registry re3data, focusing on typologies visible in their public interfaces.

The analysis demonstrated that the platforms adopt representation approaches characterized by a high level of generalization: Zenodo primarily organizes research outputs according to resource formats, whereas re3data structures repositories through broad disciplinary categories. While such approaches support interoperability and large-scale data organization, they may not adequately capture the epistemological diversity and methodological complexity that characterize SSH research data, and may affect their representation, retrieval, discoverability, and dissemination.

The findings suggest that typologies implemented in general research data infrastructures only partially represent the diversity of SSH research practices, indicating the need for more context-sensitive and epistemologically informed representation approaches. From a Knowledge Organization perspective, the study highlights that research data infrastructures are not neutral repositories of information but active representational systems whose representation choices influence what becomes visible, retrievable, and reusable within scholarly communication. This observation reinforces the importance of incorporating domain-sensitive perspectives into the design of data infrastructures, particularly when dealing with heterogeneous research communities such as the SSH.

The present study also provides a foundation for broader doctoral research investigating how SSH researchers manage and share research data. Future research will incorporate surveys and interviews with researchers and extend the analysis to discipline-specific repositories, enabling comparisons between general-purpose and domain-oriented infrastructures. Such investigations may contribute to the development of more nuanced

approaches to research data representation that better accommodate the epistemological diversity of SSH knowledge production.

AI Usage Statement

No generative artificial intelligence tools were used in the preparation of this work. Canva was used as a graphic design tool for the creation of Figure 3.

References

Borgman, Christine L. 2015. Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, MA: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press.

Bowen, Glenn A. 2009. "Document Analysis as a Qualitative Research Method." Qualitative Research Journal 9, no. 2: 27–40. doi: 10.3316/QRJ0902027.

Gualandi, Bianca, Luca Pareschi, and Silvio Peroni. 2023. "What Do We Mean by 'Data'? A Proposed Classification of Data Types in the Arts and Humanities." Journal of Documentation 79, no. 7: 51–71. doi: 10.1108/JD-07-2022-0146.

Hjørland, Birger. 2017. "Domain Analysis." Knowledge Organization 44: 436–464. doi: 10.5771/0943-7444-2017-6-436.

Huvila, Isto, and Luanne Sinnamon. 2022. "Sharing Research Design, Methods and Process Information in and out of Academia." Proceedings of the Association for Information Science and Technology 59: 132–144. https://doi.org/10.1002/pra2.611.

Huvila, Isto, and Luanne S. Sinnamon. 2024. "When Data Sharing Is an Answer and When (Often) It Is Not: Acknowledging Data-Driven, Non-Data, and Data-Decentered Cultures." Journal of the Association for Information Science and Technology 75: 1515–1530. doi: 10.1002/asi.24957.

Lin, Chi-Shiou. 2023. "Exploring Humanities Researchers' Perceptions of 'Data': A Phenomenological Approach." Proceedings of the Association for Information Science and Technology 60: 1052–1054. doi: 10.1002/pra2.942.

OECD. 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD Publishing. doi: 10.1787/9789264034020-en-fr.

Pereira, Isabella Henrique Lima, and Suellen Oliveira Milani. 2025. "Organização dos dados de pesquisa em Ciências Sociais e Humanas: uma reflexão sobre assuntos e tipos de conteúdo no Registry of Research Data Repositories." In Desafios Éticos na Organização do Conhecimento II, edited by Fernanda Ribeiro, 465–480. Porto: Faculdade de Letras da Universidade do Porto.

Poljak Bilić, Ljiljana, and Kristina Posavec. 2024. “FAIRness of Research Data in the European Humanities Landscape.” Publications 12, no. 1: 6. doi: 10.3390/publications12010006.

Toffler, Alvin. 1980. The Third Wave. New York: William Morrow. https://openlibrary.org/books/OL4422245M/The_third_wave.

Given, Lisa M., and Rebekah Willson. 2018. “Information Technology and the Humanities Scholar: Documenting Digital Research Practices.” Journal of the Association for Information Science and Technology 69: 807–819. https://doi.org/10.1002/asi.24008.

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: 377–395. doi: 10.1002/asi.10387.