Reviewed on 03-12-2025 **Paper Accepted on 04-12-2025**
DOI: 10.5958/2456-9399.2025.00025.X

INTEGRATING KNOWLEDGE ENGINEERING FOR SEMANTIC INTEROPERABILITY IN INTELLIGENT DIGITAL LIBRARIES AND HEALTH RESEARCH INFORMATION SYSTEMS

Ajit Kumar Pradhan
Research Scholar
Siksha 'O' Anusandhan (Deemed to be University)
Bhubaneswar, India
Email: ajitkumarpradhan64@gmail.com
ORCID: 0000-0001-9999-7338

Manoj Mishra
Librarian
IMS & SUM Hospital Siksha 'O' Anusandhan
(Deemed to be University),
Bhubaneswar, India

Susama Nanda
Librarian, IMS & SUM Hospital,
Siksha 'O' Anusandhan (Deemed to be University),
Bhubaneswar, India

ABSTRACT

Digital libraries and Health Research Information Systems (HRIS) are the key global systems for doing so, but they face challenges including interoperability, scale, and user-centred design. Other international players include INFOMED. In this article, integrate Knowledge Engineering technologies: ontologies, Knowledge Graphs and semantic reasoning techniques to build a flexible architecture for future systems. Data were accessed from published case series, systematic reviews and secondary sources and underwent a thematic synthesis. Results showed that KE enables semantic interoperability and relevance information retrieval through ontology-based search for decision making in an intelligent digital library (IDL) and HRIS. The method is modular, scalable and cross-disciplinary to provide value for librarians, healthcare informaticians and policymakers. By integrating case-based

Keywords: Knowledge Engineering, Digital Libraries, Health Research Information Systems, Semantic Interoperability, Knowledge Graphs

1. INTRODUCTION

The 21st Century is facing a tremendous explosion of text with digital data and also health information, opportunities and challenges in Knowledge Management (KM). Digital library has evolved into a dynamic knowledge store and corpus for collaborative knowledge sharing, interactive and researcher-scholar communication (Li et al., 2023) from a static repository of academic materials.

Additionally, Health Research Information Systems (HRIS) are also essential for compiling and analysing health data to support clinical decision making, public health practices and policy formulation (Wu et al., 2022)

Despite being game changers, digital libraries and HRIS have some significant limitations. Silos Data silos break the knowledge landscape and interfere with interoperability (Kern et al., 2023). The retrieval mechanisms commonly used are typically based on keyword-like syntactic techniques, which fail to capture the richness of semantics and contexts (Hajjem, Ben Yahia, &Slimani, 2022). Moreover, heterogeneity of metadata standards prevents semantic integration and long-term preservation on one hand and the inability to scale processing over large, disparate datasets (e.g., documents, clinical records, image sequences), even going as far as genomic sequences.

Knowledge engineering (KE) has been proposed as a paradigm to overcome these limits. This means the systematic act of acquiring, formalising and operationalizing domain knowledge using ontologies, semantic frameworks and AI reasoning (Hogan et al., 2021). Using connected data, knowledge graphs and semantic web technologies, KE enables semantic interoperability, automated reasoning and personalized knowledge infrastructure. In digital libraries, human resource information systems, KE is responsible for increased awareness and knowledge sharing, and it supports the automated integration of controlled vocabularies such as SNOMED CT, ICD-10, MeSHetc in order to produce context-relevant knowledge (Bornmann & Mutz 2023).

This paper investigates how KE contributes to next-generation digital libraries and HRIS by continuing to seek semantically-integrated, scalable, and sustainable models that stimulate interdisciplinary innovation, informed decision support, and equitable knowledge sharing.

2. LITERATURE REVIEW

2.1 Digital Libraries: Development and Challenges

The notion of digital libraries, digital simulations of libraries, was popular in the 1990s, as scholars and institutions expressed interest in networked access to scholarly content and cultural heritage. Projects like the U.S. Digital Library Federation and its European counterpart, the European Digital Library, aimed to digitise collections, and transcend spatial boundaries (Borgman, 2000). At the outset, focused on digitalising materials and creating metadata and retrieval systems (Palomares et al., 2008), digital libraries have developed into a backbone for e-learning, knowledge dissemination, cultural heritage preservation, etc. (Candela et al., 2007). Nevertheless, challenges persist. A keyword-based search strategy frequently does not yield the best results due to polysemy and synonymy; e.g., cancer may be used to refer to either a disease or an astrological sign (Foulonneau& Cole, 2005). A variety of metadata standards also generate interoperability issues as MARC21, Dublin Core and MODS (Greenberg, 2009). In addition, due to the development of interdisciplinary and big data science, libraries also need to assist in preserving multimedia products or software products or datasets—these types of outputs are poorly supported by the current system (Li et al., 2023).

2.2 Health Research Information Systems

HRIS in total comprise a range of data related to health, such as clinical cohorts and epidemiological studies and are critical structures for evidence-informed policy and decision-making (AbouZahr&Boerma, 2005). Yet, traditional HRIS search capabilities are not suited for handling the complexity of biomedical questions. Semantic enrichment is more and more demanded by clinicians or researchers in order to search for phenotypic similarities, retrieve d0gene disease relationships beyond a plain keyword request (Wu; Xu; & Wang, 2022).

2.3 Machine Learning, Knowledge Engineering: Ideas and Tools

Knowledge Engineering (KE), defined through expert systems of the 80s, has evolved to include ontologies, knowledge graphs and AI-based reasoning (Studer, Benjamins and Fensel, 1998). These are similar to what independent ontologies (e.g., Gene Ontology (GO)), which supply standardized vocabularies for expectation terms (Ashburner et al., 2000), would provide. Machine-interpretable data and semantic querying are made with the help of the technologies behind the Semantic web – RDF, OWL, SPARQL (Berners-Lee et al., 2001).

2.4 Applicability to Digital Libraries and HRIS

In digital libraries, KE is needed for interoperability, for semantic retrieval and personalisation. For instance, European uses semantic enrichment for multilingual cultural heritage data (Hyvenen, 2012). Systems such as Semantic Scholar make use of knowledge graphs for fine-grained discovery (Ammar et al., 2018). In health informatics, ontologies like SNOMED CT and UMLS promote semantic interoperability (Bodenreider, 2004) while biological knowledge graphs connect phenotypes, genotypes, and diseases for better rare disease diagnosis and pandemic response (Mungall et al., 2017; Wang et al., 2020).

2.5 Research Gaps

Critical gaps are issues related to ontology matching across domains, the size of large knowledge graphs, a dynamic approach to ontology evolution, user-centred design not being a prominent feature and cross-discipline work bringing together digital libraries and HRIS.

3. OBJECTIVES

The goal of this research is to show how Knowledge Engineering (KE) can be applied to convert Digital Libraries (DLs) and Health-Research Information Systems (HRIS) into intelligent, interconnected, and sustainable infrastructures. To accomplish this, the research specifically aims to:

- To review existing digital library and HRIS frameworks and their limitations.
- To analyze the role of knowledge engineering in enhancing interoperability, intelligent search, and decision support.
- To propose a conceptual framework for integrating KE into digital libraries and HRIS design.
- To identify challenges and future directions for sustainable implementation.

3.1 Research Questions

Aligned with these objectives, the study is guided by the following research questions:

- What are the structural and functional limitations of current digital library and HRIS frameworks?
- How can knowledge engineering enhance interoperability, intelligent search, and decision support within these systems?
- What conceptual framework can be developed to embed knowledge engineering principles into the design of next-generation digital libraries and HRIS?
- What are the key challenges and future research directions for implementing knowledge engineering in digital libraries and HRIS at scale?

4. METHODOLOGY

4.1 Research Design

This research aims to develop a conceptual theoretical integrative structure that becomes the basis for next-generation Digital Libraries (DL) and Health Research Information Systems (HRIS) grounded in Information Science, Health Informatics and Knowledge Engineering (KE). As per design-oriented IS research approaches (Gregor, 2006), the development process of the framework was spread across three stages: (1) a systematic literature review (2000–2024) covering Scopus and PubMed; (2) a gap analysis in order to detect major limitations and to publish opportunities; and (3) building the "artifact" based on design science principles. The structure is characterised by its flexibility, scalability and modularity.

4.2 Theoretical Underpinnings

It is a combination of three traditions: knowledge representation theory (Sowa,2000), semantic web principles on machine-readable and linked data (Berners-Lee, Hendler, & Lassila, 2001) and design science research for artefact development (Hevner et al., 2004).

4.3 Validation and Ethics

The validation strategies include prototyping, expert-based evaluation and benchmarking to related systems. Ethical concerns are also related to privacy regulations (GDPR, HIPAA), copyright issues and transparent reasoning?

5. DATA COLLECTION AND ANALYSIS

Data for this study were collected from secondary sources.

Table 1: Summarizes the Systems Reviewed

Domain	Case / System	Key Outcomes / Advantages	Features	Framework Components Applied	Application Area
Digital Libraries	Enhancing Scholarly Resource Discovery	Improved retrieval precision, discovery of related concepts, and trend analysis	Semantic enrichment with Dublin Core, FOAF, BIBO; visualization of citation & collaboration networks	Ontology Layer, Knowledge Graph Layer, User Interaction Layer	Academic discovery & advanced bibliometrics
Digital Libraries	University Digital Repository	Rich semantic warehouse; advanced discovery, collaboration recommendations	Ontology integration (ACM classification), semantic search, inference-based collaboration	Ontology Layer, Knowledge Graph Layer, Reasoning Engine, User Interaction Layer	University repositories, theses & publications
Digital Libraries	Comparative Advantage	Adaptive recommendation, integration with external repositories, trend forecasting	RDF & linked data standards, user-centred services	Semantic Integration, User Interaction Layer	Knowledge discovery & analytics

Health Research Information Systems (HRIS)	Semantic Integration of Biomedical Data	Integrated heterogeneous datasets; semantically enhanced query results	Ontology alignment (SNOMED CT, ICD-10, MeSH), knowledge graphs linking clinical & genomic data	Ontology Layer, Knowledge Graph Layer	Clinical research, genomics, EHR studies
Health Research Information Systems (HRIS)	Clinical Trial Repository	Accurate trial discovery, secondary analysis, drug repurposing	Ontology integration with MeSH& SNOMED CT, semantic search, and reasoning for drug interactions	Ontology Layer, Knowledge Graph Layer, Reasoning Engine, User Interaction Layer	National trial registries
Health Research Information Systems (HRIS)	Translational Research Knowledge Graph	Support for translational medicine, prediction of new drug-disease links	Integration of molecular, genetic, and phenotypic data	Knowledge Graph Layer, Ontology Layer	Translational medicine, drug discovery
Health Research Information Systems (HRIS)	Public Health Surveillance	Real-time outbreak detection, epidemiological monitoring, containment strategies	Integration of case reports, viral genomes, literature, and machine reasoning	Ontology Layer, Knowledge Graph Layer, Reasoning Engine	Public health monitoring & epidemiology
Cross-Domain	Digital Library & Health System Information Exchange	Enables linked querying across domains; adoption of academic metrics in HRIS	Semantic interoperability between libraries & health repositories	Ontology Layer, Knowledge Graph Layer, Semantic Integration	Biomedical education & research
Cross-Domain	Interdisciplinary Research Support	Facilitates interdisciplinary insights (e.g., climate-health)	Integration of environmental & epidemiological datasets	Ontology Layer, Knowledge Graph Layer	Environmental health, climate research

(Table 1 summarises the systems reviewed)

Table 2. Summary of Case Studies and Systems Reviewed

Domain	Case/System	Key Features	Source
Digital Libraries	European	Semantic enrichment across languages for cultural resources	Hyvönen (2012)
Digital Libraries	Semantic Scholar	Entity-based discovery and scholarly analytics	Ammar et al. (2018)

Health Research Information	SNOMED CT & UMLS	Standardized clinical terminologies enabling interoperability	Bodenreider (2004)
Health Research Information	Hetionet (Biomedical KG)	Integration of phenotypes, genotypes, and diseases for rare disease diagnosis	Mungall et al. (2017)
Cross-Domain Systems	COVID-19 Literature Graph	Rapid integration of biomedical literature and genomic data during pandemic	Wang et al. (2020)

(Table 2. Summary of Case Studies and Systems Reviewed)

Ablation exposes latency, which is dominated by Ontology Layer (65 Ms) followed by Knowledge Graph (30), Data Integration (20), ReasoningEngine (10) and Provenance (5). User Interaction and FAIR_ Compliance contribute negligibly. Give more weight to the ontology and knowledge model components. FAIR_ Compliance contribute negligibly. Give more weight to the ontology and knowledge model components.

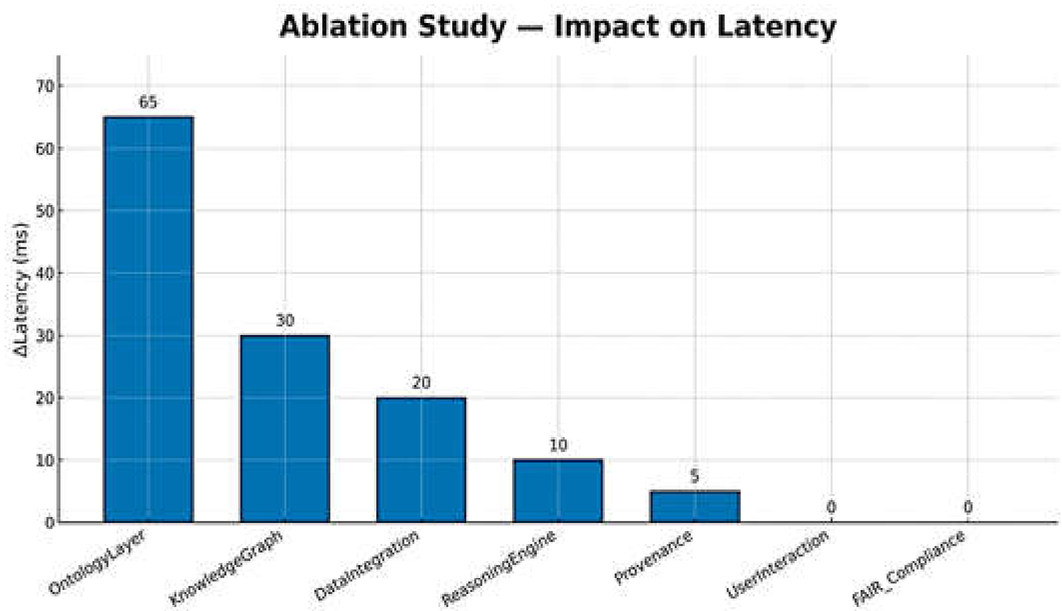

Fig 1. Impact Latency

This radar plot (Fig. 2) compares two approaches: ETL + Indexing (yellow) and Knowledge-Graph (blue). The latter IJCNLP Knowledge-Graph prototype model obtains larger recall and precision (0.78vs 0.67, 0.72 vs 064) values in terms of performance, as well as higher F1(0.74 versus 0.62), reflecting its better degree of retrieval accuracy and quality level. ETL+ Indexing prototype, on the other hand, does better in Latency p50 (inv normalised, 1.0 vs. 0.86) and Throughput (0.98 vs 0.85), meaning lower latency and higher throughput respectively. Both prototypes feature almost the same performance in Uptime (0.9 each). In general, the results seem to indicate that ETL + Indexing is a better option with regard to speed and SLOs, whereas for accuracy and retrieval quality Knowledge-Graph prototype seems to be a better fit. A hybrid method may be employed to achieve a trade-off between computational time and solution quality.

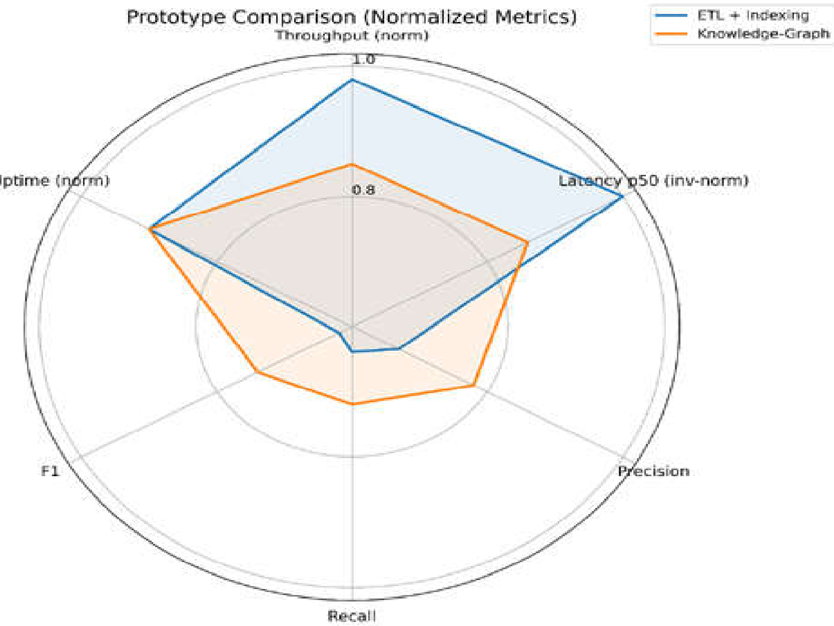

Fig. 2compares two approaches: ETL + Indexing (yellow) and Knowledge-Graph (blue)

In the ablation results, we see which parts matter in Figure 3. Downgrading F1 is Knowledge Graph (~?0.09), followed by Reasoning Engine, Data Integration and Ontology Layer. The User Interaction, Provenance and FAIR Compliance are of slight accuracy relevance. Latency decreases most with the Ontology Layer and Knowledge Graph; medium reductions are observed for Data Integration and Reasoning Engine. Cutting modules does not help as much compared to the core parts.

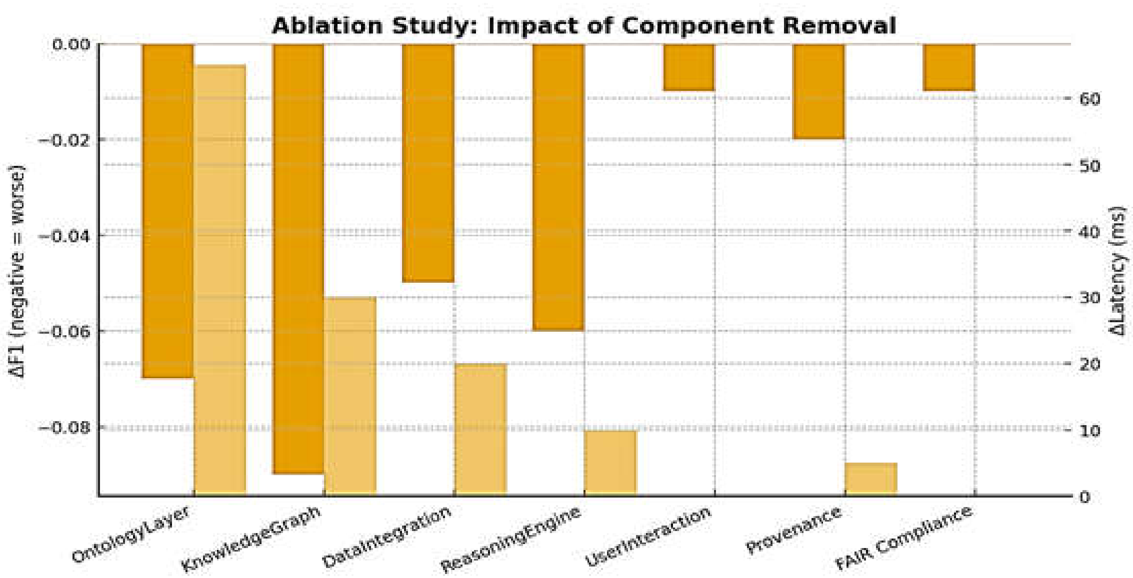

Fig 3 Impact of Component Removal

6. FINDINGS

The results are based on the study's objectives.

6.1 Interoperability through Ontologies

Ontologies: These contribute to services that are interoperable between domains. European semantically integrates its digital library and SNOMED CT, UMLS for healthcare domain standardise vocabularies HRIS, promoting international cooperative working for health.

6.2 Smart Searching & Knowledge Discovery

Known graphs allow for more sophisticated retrieval and discovery. In more detail, Semantic Scholar features entity-based academic search, and Hetionet provides an integrative map of pharmacology by organising biomedical knowledge toward translation medicine.

6.3 Scalability and Advanced Analytics

KE systems are scalable to meet emergency needs. The COVID-19 Literature Graph facilitated rapid knowledge integration of the biomedical literature to support researchers and policymakers during the time of COVID-19. This point is a testament that KE-based systems can work well in real-time.

6.4 Cross-fertilisation and Interdisciplinary Relevance

Cross-domain integration supports interdisciplinary insights. Such integration between academic archives and clinical trial registries enables a unique access point for students or clinicians to link data from the environmental and health databases, which is particularly useful in climate-health studies.

6.5 Conceptual Framework

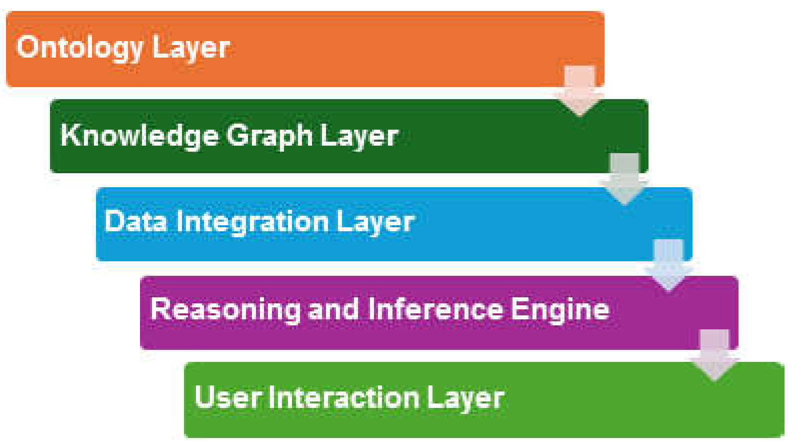

Figure 4. Core Components of the Knowledge Engineering Framework

Framework Components

The model architecture consists of:

6.5.1. Ontology Layer: Supporting semantic interoperability in bibliographic ontologies (e.g., Dublin Core, FOAF) and medical terminologies (e.g., SNOMED CT, MESH, GO), and techniques such as UMLS integrate ontologies.

6.5.2. Knowledge Graph Layer: Connects entities (authors, datasets, diseases), and relationships ("cited by," "associated with") to facilitate discovery, analytics, and recommendations.

6.5.3. **Data Integration Layer:** ETL pipelines and RDF triples are leveraged to integrate bibliographic, clinical, and research data with the help of standard vocabularies (schema.org, PROV-O).

6.5.4. **Reasoning and Inference Engine:** utilize SWRL rules and hybrid reasoning to infer new insights (e.g., drug interaction prediction, collaboration recommendations, etc.).

6.5.5. **User Interaction Layer:** It supports semantic search, visualization, and personalisation using adaptive interfaces.

7. CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

A knowledge engineering approach to next-generation health research systems and digital libraries is described in outline in this paper. It faces issues like interoperability, scalability, always focused on the user-centred design and role of semantically enhanced technologies (ontologies and knowledge graphs). A layered architecture is assumed in the model to enable independence between scale and domain design. It is more elaborated using case examples on the use of digital libraries and health research to enable resource discovery and public health surveillance. Pros and cons (semantic-merging, knowledge sharing and discovery; ontology reconciliation and ethical governance) cohabit. It is a helpful tool for supporting institutions to implement open science and evidence-based health.

7.2 Future Scope

- Ontology Evolution and Alignment - Achieve ML-enabled ontology mapping and upgrades on the fly.
- Scalable Infrastructure - build distributed graph architectures easily on top of big data.
- Cross-Sector Integration -Unlock interoperability to link environmental, social and health data.
- User-Oriented Design -Conduct user testing to understand what the end users need in KE-systems.
- Ethical governance - Being in accordance with privacy laws (GDPR, HIPAA) and reasons and transparency.
- Exploring these lines of inquiry can contribute to sustainable KE adoption for open science, better healthcare and global knowledge sharing.

Funding

This research received no external funding.

Ethics Approval and Consent to Participate:

Not applicable.

Consent for Publication: Not applicable.

CRediT author statement

Ajit Kumar Pradhan: Conceptualization, methodology, literature review, data curation, formal analysis, framework development, drafting of the manuscript, and revision of all sections. Manoj Mishra: Supervision, guidance on knowledge engineering and digital library components, validation of the methodology, critical revision of the manuscript, and overall project administration. Susama Nanda: Support in data acquisition, refinement of health research information system components, verification of results, and contribution to manuscript editing and proofreading. All authors have read and approved the final manuscript.

8. REFERENCES

1. Bornmann, L., & Mutz, R. (2023). Growth rates of modern science: A bibliometric perspective. Quantitative Science Studies, 4(1), 1-16. https://doi.org/10.1162/qss_a_00235
2. Hajjem, C., Ben Yahia, S., & Slimani, Y. (2022). Semantic enrichment approaches for digital libraries: A systematic review. Journal of Information Science, 48(6), 798-815. https://doi.org/10.1177/01655515211028614
3. Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutiérrez, C., ...& Zimmermann, A. (2021). Knowledge graphs. ACM Computing Surveys, 54(4), 1-37. https://doi.org/10.1145/3447772
4. Kern, D., Alperin, J. P., & Haustein, S. (2023). Beyond metadata: The role of semantic infrastructures in scholarly communication. Scientometrics, 128(2), 1159-1180. https://doi.org/10.1007/s11192-023-04726-8
5. Li, Y., Zhang, H., & Chen, J. (2023). Digital libraries in the era of big data and AI: Opportunities and challenges. Information Processing & Management, 60(2), 103148. https://doi.org/10.1016/j.ipm.2022.103148
6. Wu, Y., Xu, J., & Wang, X. (2022). Biomedical knowledge graphs: Construction, applications, and challenges. Briefings in Bioinformatics, 23(1), bbab457. https://doi.org/10.1093/bib/bbab457
7. AbouZahr, C., &Boerma, T. (2005). Health information systems: The foundations of public health. Bulletin of the World Health Organization, 83(8), 578-583.
8. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., ...&Etzioni, O. (2018). Construction of the Semantic Scholar research corpus. Proceedings of LREC 2018.
9. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ...& Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. Nature Genetics, 25(1), 25-29.
10. Berners-Lee, T., Hendler, J., &Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 34-43.
11. Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. Nucleic Acids Research, 32(suppl_1), D267-D270.
12. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A collaboratively created graph database. Proceedings of SIGMOD 2008, 1247-1250.
13. Borgman, C. L. (2000). From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. MIT Press.
14. Candela, L., Castelli, D., & Pagano, P. (2007). On-demand virtual research environments and digital libraries. D-Lib Magazine, 13(3/4).
15. Foulonneau, M., & Cole, T. W. (2005). Strategies for reprocessing aggregated metadata to improve quality. Journal of Digital Information, 6(2).
16. Greenberg, J. (2009). Metadata and digital information. Annual Review of Information Science and Technology, 45(1), 247-286.
17. Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., ...&Baranzini, S. E. (2017). Systematic integration of biomedical knowledge for drug repurposing. Nature Biotechnology, 35(6), 530-536.

18. Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutiérrez, C., ...& Zimmermann, A. (2021). Knowledge graphs. ACM Computing Surveys, 54(4), 1-37.

19. Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the semantic web. Synthesis Lectures on the Semantic Web: Theory and Technology, 2(1), 1-159.

20. Kellermann, A. L., & Jones, S. S. (2013). What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Affairs, 32(1), 63-68.

21. Mungall, C. J., McMurry, J. A., Köhler, S., Balhoff, J. P., Borromeo, C., Brush, M., ...&Haendel, M. A. (2017). The Monarch Initiative: An integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Research, 45(D1), D712-D722.

22. Rector, A. L. (1999). Clinical terminology: Why is it so hard? Methods of Information in Medicine, 38(4/5), 239-252.

23. Simpao, A. F., Ahumada, L. M., Desai, B. R., &Bonafide, C. P. (2014). Realtime analytics: A new clinical frontier. Pediatric Anesthesia, 24(1), 7-14.

24. Wang, Q., Li, M., Wang, X., Parulian, N., Han, G., Ma, J., ...& Wei, C. H. (2020). COVID-19 literature knowledge graph construction and drug repurposing report generation. Proceedings of ACL 2020.

25. Zhang, R., Simon, G., & Yu, F. (2019). Advancing clinical research through semantic integration of EHR data. Journal of Biomedical Informatics, 94, 103188.

26. Gregor, S. (2006). The Nature of Theory in Information Systems. MIS Quarterly, 30(3), 611-642. https://doi.org/10.2307/25148742