GRADIVA

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

# Linked Data Technologies in Scholarly Communication: A Bibliometric and Thematic Analysis (2000–2024)

Dr. Amreen Taj[1]*, Shashikumar Jawadagi[2] & Dr. Binsha Pappachan C[3]

1.Assistant Librarian, Yenepoya Central Library, Yenepoya (Deemed to be University), Mangaluru, Karnataka. *Corresponding Author Email: czkamreentaj@gmail.com, ORCID: https://orcid.org/0000-0001-5624-3931
2.Professor, Department of Medical Surgical Nursing, Yenepoya Nursing College, Mangalore. E-mail: shashikumar@yenepoya.edu.in
3.Professor/HOD, Department of Mental Health Nursing, Yenepoya Nursing College, Mangalore. E-mail: binshadennis@yenepoya.edu.in

## Abstract

This study presents a bibliometric and thematic analysis of Linked Data research from 2000 to 2024, focusing on applications in scholarly communication, libraries, and digital knowledge systems. Using a TITLE-based search in Scopus. Findings reveal sustained growth in Linked Data scholarship, peaking between 2010 and 2016. Libraries emerge as primary adopters for bibliographic control, authority management, and knowledge graph integration, although challenges remain in legacy data conversion, vocabulary alignment, and standardization. The results provide a comprehensive overview of Linked Data's evolution and its role in advancing open, interoperable, and machine-actionable scholarly infrastructures.

**Keywords:** *Open Knowledge, Digital Inclusion, FAIR Principles, Metadata Interoperability, Library Science Applications.*

## 1. INTRODUCTION

The growing relevance of structured and machine-readable data has positioned Linked Data as a important area of research. As an extension of the Semantic Web, Linked Data seeks to build a network of interconnected and interoperable datasets using foundational technologies such as Uniform Resource Identifiers (URIs), the Resource Description Framework (RDF), and SPARQL query language (Gunaratna et al., 2014; Hogan, 2014; Mountantonakis & Tzitzikas, 2019; Quarati & Albertoni, 2024). Tim Berners-Lee first envisioned the integration of documents and data in 1989 and, along with Ora Lassila, formally introduced the Semantic Web in 2001. By 2006, Berners-Lee had articulated the core design principles of Linked Data, providing a framework that has since guided its implementation, particularly within library science, digital repositories, and knowledge management (Bizer et al., 2009, 2023).

With the steady expansion of Linked Data scholarship, bibliometric and scientometric methods have become instrumental in evaluating its academic influence. Bibliometric analysis quantitatively assesses publication outputs, citation impact, and research themes, while scientometric techniques delve deeper to map conceptual networks and detect emerging research directions. These complementary approaches offer a holistic view of the discipline's intellectual development, revealing major contributors, collaborative networks, and thematic shifts.

Celebrating its tenth anniversary in 2023, the San Francisco Declaration on Research Assessment (DORA) remains committed to promoting fairer and more meaningful approaches to research evaluation. In this context, bibliometric studies play a vital role in assessing

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

research output effectively. Moving beyond journal-based metrics like the Impact Factor, DORA promotes diverse qualitative and quantitative criteria that better capture the full value of scholarly contributions (Gingras, 2016).

### 1. 2. Study Objectives

- To assess the publication and citation growth patterns in Linked Data (LD) research from 2000 to 2024.
- To identify the most productive and influential contributors, including authors, institutions, and sources shaping LD scholarship.
- To explore the conceptual and thematic evolution of LD research through keyword co-occurrence, factorial analysis, and thematic mapping.
- To map international and inter-institutional collaboration networks in LD research using co-authorship and country-level data.
- To examine the disciplinary integration of LD with library and information science, computer science, and open science domains.
- To analyze the impact of LD research on metadata interoperability, bibliographic control, and FAIR data implementation in scholarly communication and library systems.

### 1.3 Research Questions

1. What are the publication and citation trends in Linked Data research between 2000 and 2024?
2. Who are the most influential authors, institutions, and publication sources in the LD research landscape?
3. What are the dominant, emerging, and declining research themes in LD, as revealed through keyword and thematic analysis?
4. How is scholarly collaboration structured geographically and institutionally in LD research?
5. What are the conceptual structures and topic clusters driving LD research across disciplines?
6. How has Linked Data research contributed to advancing FAIR principles, metadata interoperability, and bibliographic control, particularly in the library and information science domain?

## 2. LITERATURE REVIEW

### 2.1 Evolution of Linked Data Research

The concept of Linked Data (LD) originated from Tim Berners-Lee’s 2006 proposal, which emphasized the use of URIs(Uniform Resource Identifiers), , RDF (Resource Description Framework), and SPARQL Protocol and RDF Query Language (SPARQL) to create an interconnected web of structured, machine-readable data (Bizer et al., 2023). This vision marked a shift from unstructured documents to a Semantic Web, enabling data sharing and interoperability across domains (Hitzler & Janowicz, 2013).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

One of the earliest large-scale implementations was the Linked Open Data (LOD) Cloud, introduced in 2007, which visualized the relationships between publicly available RDF datasets. This initiative catalyzed cross-domain adoption, including early exploratory projects in libraries and cultural heritage institutions (Bizer et al., 2023).

The library and archives community began engaging with Linked Data through pilot studies and experimental frameworks. In 2010, institutions like the Library of Congress, British Library, and Europeana launched initial prototypes to test the viability of LD for bibliographic metadata (Hallo et al., 2016). These efforts were further supported by the W3C Library Linked Data Incubator Group (2011), which advocated for metadata standardization and encouraged best practices for semantic data modeling in libraries (W3C, 2011).

A major milestone was the launch of BIBFRAME (Bibliographic Framework Initiative) by the Library of Congress in 2015. Designed to replace the MARC format, BIBFRAME aimed to serve as a Linked Data model for bibliographic description (Fortier et al., 2022; Jin et al., 2016). Although still under development and pilot testing in many institutions, BIBFRAME represents an important shift toward integrating semantic data in library catalogs.

Complementary projects such as LD4L (Linked Data for Libraries) and LD4P (Linked Data for Production), primarily funded by the Mellon Foundation, have focused on creating Linked Data pipelines, ontologies, and tools to support metadata enrichment and interinstitutional collaboration (Mountantonakis & Tzitzikas, 2019). While some outputs of these initiatives remain experimental, others have been adopted in academic library environments.

The Europeana Data Model (EDM), developed to unify cultural heritage metadata across Europe, also exemplifies the practical implementation of semantic models. EDM leverages RDF and Linked Data principles to facilitate integration of heterogeneous data sources (Doerr et al., 2010). Similar efforts include the Smithsonian Libraries' semantic collections, the Linked Logainm project in Ireland (which links to DBpedia and Geonames), and the Digital Public Library of America (DPLA), all of which demonstrate the expanding application of LD in enhancing metadata discoverability and cross-domain linking (Gaitanou et al., 2024).

However, many of these projects remain semi-operational, prototype-based, or in transition phases rather than fully deployed systems. This distinction is critical when evaluating the maturity of LD technologies in libraries.

Despite progress, significant challenges persist, such as inconsistent data quality, lack of standardization, and technical complexity associated with integrating legacy metadata into RDF frameworks (Zaveri et al., 2015; Harron et al., 2017). These barriers often limit the scalability and interoperability of LD implementations.

Looking forward, the field is shifting toward more robust approaches, such as incorporating AI, Big Data analytics, and automated validation tools to improve data quality, user interaction, and dynamic metadata enrichment (Hyvönen et al., 2025). These advancements are crucial for realizing the long-term vision of semantically rich, interconnected, and reusable data infrastructures in libraries.

This conceptual background informs the present bibliometric study, which maps research trends, key contributors, and evolving themes in Linked Data applications within library and information science. The empirical findings are situated within this evolving technological and conceptual landscape.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

## 2.2 Applications of Linked Data in Libraries and Information Science

Linked Data has significantly transformed bibliographic metadata management by enabling machine-readable, interconnected, and structured data in library systems. Traditional metadata formats such as MARC (Machine-Readable Cataloging) were designed for print-based catalogs and primarily human interpretation. However, with the rise of Linked Data and Semantic Web technologies, metadata is now more interoperable and reusable across multiple platforms (Alemu et al., 2012).

The BIBFRAME (Bibliographic Framework) model, developed by the Library of Congress, serves as a Linked Data replacement for MARC and facilitates metadata interoperability by leveraging RDF (Resource Description Framework) and URIs (Uniform Resource Identifiers) to interconnect bibliographic entities (Cole et al., 2013; Kroeger, 2013; McCallum, 2017). This transformation allows libraries to link their bibliographic records to external datasets, such as Wikidata, DBpedia, and VIAF (Virtual International Authority File), enhancing the discoverability and contextual richness of bibliographic resources (Gaitanou et al., 2022).

Efforts to integrate Linked Data with existing cataloging standards have led to the development of hybrid models that bridge legacy metadata formats with Linked Data principles. The Functional Requirements for Bibliographic Records (FRBR) model and its successor, IFLA LRM (Library Reference Model), align closely with Linked Data by structuring bibliographic entities into a hierarchical, relationship-based framework (Candela et al., 2022).

One major challenge in transitioning from MARC to Linked Data has been the complexity of converting legacy records into structured, linked metadata. To address this, projects such as MARC-to-RDF conversion tools have been developed, enabling libraries to transform their records into RDF triples for integration into the Semantic Web (Alemu et al., 2012). Similarly, BIBFRAME is designed to be backward-compatible with MARC, allowing for a gradual adoption of Linked Data principles while maintaining existing cataloging workflows (Gaitanou et al., 2022).

The integration of Linked Data with FRBR-based models has also facilitated the creation of entity-based catalogs, where works, authors, and subjects are uniquely identified and interconnected. This approach allows for more precise search results, better resource recommendations, and improved bibliographic relationships across library collections (Heng et al., 2024).

Digital libraries and institutional repositories have adopted Linked Data technologies to enhance metadata interoperability, data discoverability, and knowledge organization. The Europeana Data Model (EDM) is a widely adopted Linked Data framework that integrates cultural heritage data from multiple institutions, facilitating cross-repository search and discovery (Capurro & Plets, 2020; Dutta et al., 2025; Peroni et al., 2013).

In academic and research libraries, Linked Data has enabled richer metadata representation and semantic search capabilities in institutional repositories. The Digital Public Library of America (DPLA) employs Linked Data to interconnect metadata records from diverse institutions, enhancing access to digital collections (Warraich & Rorissa, 2024). Additionally, university libraries have implemented RDF-based repositories to improve metadata interoperability and link scholarly content to external knowledge graphs (Shah et al., 2025).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

However, the adoption of Linked Data in digital libraries is not without challenges. Studies highlight technical barriers, lack of expertise, and concerns over data quality and sustainability as key impediments to widespread Linked Data adoption (Saleem et al., 2018). To address these challenges, initiatives such as Linked Data for Production (LD4P) and Sinopia (a Linked Data editor for bibliographic metadata) have been developed to support libraries in transitioning to Linked Data-based cataloging (Heng et al., 2024).

### 2.3 Linked Data and Metadata Interoperability

Metadata interoperability is a critical requirement in digital libraries, archives, and repositories, enabling data integration and cross-system resource discovery. Linked Data (LD) plays a significant role in metadata crosswalks and schema mapping, facilitating seamless data exchange between different metadata standards. Traditional metadata schemas such as MARC (Machine-Readable Cataloging), Dublin Core (DC), METS (Metadata Encoding and Transmission Standard), MODS (Metadata Object Description Schema), and EAD (Encoded Archival Description) often lack direct interoperability due to structural differences (Huang & Qin, 2024). Linked Data provides a semantic bridge through the use of RDF (Resource Description Framework) and URIs (Uniform Resource Identifiers), ensuring that metadata elements from different schemas are aligned and interconnected (Thalhath et al., 2025).

A significant approach to metadata crosswalks is the use of application profiles, which define customized vocabularies to align metadata fields across diverse standards. For instance, the Dublin Core Application Profile (DCAP) enables interoperability between different metadata schemas, allowing digital libraries to map their metadata to global standards (Thalhath et al., 2025). Additionally, the BIBFRAME model developed by the Library of Congress serves as a Linked Data alternative to MARC, providing a structured, RDF-based framework that enables semantic connections between bibliographic entities (Li et al., 2025).

The NAISC-L interlinking framework has also been implemented in libraries, archives, and museums (LAMs) to create provenance-rich metadata interlinks, enhancing discoverability and alignment between disparate metadata schemas (McKenna et al., 2022). Such frameworks assist institutions in automatically linking their records with external knowledge bases such as DBpedia, Wikidata, and VIAF (Virtual International Authority File), ensuring that metadata elements are mapped meaningfully across datasets (McKenna et al., 2022).

Despite the benefits of Linked Data in metadata crosswalks and schema mapping, several challenges persist. One of the key difficulties is the inconsistent use of metadata vocabularies across institutions. While Linked Data facilitates interoperability, different organizations often use non-standard metadata fields, making schema mapping complex and requiring significant manual intervention (Hosseini et al., 2025).

Another major challenge is legacy metadata conversion. Many institutions still rely on MARC-based systems, and converting MARC records into Linked Data models such as BIBFRAME or RDF triples requires automated mapping tools and extensive data cleaning (Li et al., 2025). The lack of standardized conversion methodologies results in incomplete or incorrect mappings, reducing data accuracy and usability.

Moreover, semantic inconsistencies in metadata descriptions pose a challenge. Different metadata standards define entities and relationships differently, leading to ambiguities when integrating datasets. For example, a single bibliographic work may have multiple metadata representations in FRBR, BIBFRAME, and Dublin Core, making precise entity alignment difficult (Rezník et al., 2022).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

Technical challenges such as scalability, real-time updating of linked datasets, and metadata validation also hinder effective interoperability. Schema.org and FAIR Linked Data principles have been proposed as solutions, promoting the use of structured, machine-actionable metadata that aligns with Semantic Web standards (Frey & Hellmann, 2021). However, adoption remains uneven, particularly among smaller institutions with limited technical resources (Singh & Maurya, 2025).

Efforts to address these challenges include the development of metadata validation tools such as YAMA (Yet Another Metadata Application Profile), which supports semantic validation and mapping of non-RDF metadata to RDF (Thalhath et al., 2025). Additionally, machine learning techniques are being explored to automate metadata reconciliation and schema alignment, reducing the manual effort required for crosswalk development (Rezník et al., 2022).

Lauscher et al. (2018) emphasize the importance of open citation data in the scientific ecosystem, highlighting the role of Linked Open Citation Databases in promoting transparent and accessible scholarly communication. Their study demonstrates the feasibility of integrating linked data technologies into library workflows to improve citation extraction and curation, supporting the open science movement. Wu and Ye (2021) delve into ontology-based bibliometric research and identify semantic web and gene ontology as major clusters in ontology studies. Their bibliometric analysis across decades shows how ontology, as a subset of Linked Data, supports knowledge discovery and conceptual mapping in artificial intelligence and biomedical contexts.

A significant shift in thematic evolution is observed in Hosseini et al. (2025), who analyze the development and maturity of co-word thematic clusters in the field of linked data. Their co-word analysis, based on social network analysis and hierarchical clustering, shows that terms like 'natural language processing,' 'semantic tools,' and 'deep learning' are now central to Linked Data research, indicating a merging with computational disciplines. Ullah et al. (2018) provide a holistic review of the use of Linked Open Data in cataloging, showing a trend among libraries toward adopting frameworks such as BIBFRAME 2.0 and integrating user-generated content for metadata enrichment. Their work bridges the gap between traditional bibliographic practices and modern linked data environments.

Wahid et al. (2018) map challenges in cataloging within Linked Data environments. They reveal issues such as metadata inconsistency, technological complexity, and limited vocabulary support, suggesting that while adoption is increasing, significant institutional and technical hurdles remain. Lytras et al. (2019) shift the focus to smart cities and the personalization of library services. Their research shows how bibliometric networks and advanced analytics, including semantic annotation and research profiling, can enhance smart urban library systems and personalized knowledge delivery.

The integration of health information systems into the Linked Data paradigm is explored by the 2024 study on health information management, which highlights the interoperability of electronic health records and semantic data linkages as a growing application area for linked data technologies.

### 2.4 Bibliographic and Authority Control Using Linked Data

The integration of Linked Data into bibliographic and authority control has significantly improved metadata consistency, resource discoverability, and entity disambiguation. Authority records, which establish uniform naming conventions and standardized metadata descriptions,

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

benefit from Linked Data's ability to interlink authoritative datasets across different cataloging systems, enhancing name disambiguation and subject classification (Zhu, 2019).

One of the major advancements in authority control has been the adoption of the Virtual International Authority File (VIAF), which aggregates multiple national authority files into a single, interconnected dataset. VIAF plays a crucial role in reducing inconsistencies in name authority control, enabling multilingual support, and linking authorship records from different national libraries (Leiva-Mederos et al., 2013). Similarly, the International Standard Name Identifier (ISNI) assigns unique identifiers to authors, contributors, and publishers, ensuring that individuals with similar names are correctly distinguished across bibliographic databases (Wiederhold & Reeve, 2021).

The role of Linked Data in name disambiguation is particularly evident in Wikidata and the Library of Congress Linked Data Service, both of which provide structured metadata that uniquely identifies and interconnects authors, publishers, and other bibliographic entities. These platforms allow machine-readable access to authority records, enhancing the precision of bibliographic records and reducing duplication errors (Hakimov et al., 2012). Additionally, authority control in digital repositories has been strengthened through linked authority files, ensuring consistency in bibliographic records across different institutions (Myntti & Cothran, 2013).

The integration of Linked Data into subject classification has enhanced library catalogs' ability to link resources semantically, improving search precision and relevance. Library of Congress Subject Headings (LCSH) and other controlled vocabularies, such as the Faceted Application of Subject Terminology (FAST), are now linked to external knowledge bases, allowing users to explore related subjects across multiple datasets (Kamal & Golub, 2025). Furthermore, Resource Description and Access (RDA) frameworks have embraced Linked Data principles to improve bibliographic control in digital environments. The transition from MARC to BIBFRAME is a key example, as BIBFRAME is specifically designed to leverage Linked Data technologies, thereby enhancing interoperability, accessibility, and resource discovery (Danskin, 2013).

Despite these advancements, several challenges remain in implementing Linked Data for authority control. Variability in metadata schemas, differences in cataloging standards across institutions, and the need for broader adoption of Linked Data tools pose barriers to seamless integration (Zhu, 2019). However, ongoing advancements in Semantic Web technologies and international collaborations continue to improve the quality, reliability, and utility of Linked Data in bibliographic and authority control (Wiederhold & Reeve, 2021).

### 2.5 Knowledge Graphs and Linked Data

The use of Knowledge Graphs (KGs) in library science has expanded significantly with the evolution of Linked Data (LD) technologies. Knowledge Graphs represent structured relationships between entities, enabling more efficient information retrieval, semantic enrichment, and knowledge discovery (Haslhofer et al., 2019). The concept of Knowledge Graphs gained prominence with Google's Knowledge Graph in 2012, which enhanced search results by linking entities based on structured metadata (Jia, 2020). Libraries and digital repositories have since adopted KG-based metadata models to improve resource discoverability and interoperability across systems (Clark et al., 2022).

Libraries traditionally relied on controlled vocabularies, ontologies, and classification systems such as the Dewey Decimal Classification (DDC), Library of Congress Subject

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

Headings (LCSH), and MARC metadata. With Linked Data and Knowledge Graphs, these systems have transformed into interconnected, machine-readable datasets, allowing for semantic linking of resources across institutions (Haller et al., 2020). The National Science Library of China has implemented an academic Knowledge Graph to enhance bibliographic data integration, showcasing the potential of KGs in structured metadata management (Zhang, 2019).

Linked Data serves as the foundation for Knowledge Graphs, enabling semantic enrichment and contextual linking of library resources. Through RDF (Resource Description Framework) and SPARQL (a query language for Linked Data), libraries can connect metadata elements across different repositories, enhancing knowledge representation and automated reasoning (Jia, 2020).

One of the major implementations of Linked Data in Knowledge Graphs is Wikidata, which acts as a global hub for structured bibliographic and authority metadata. Libraries contribute to Wikidata by linking their catalogs with external sources, creating a more integrated and discoverable knowledge ecosystem (Clark et al., 2022). The Library of Congress Linked Data Service and Europeana Data Model (EDM) further illustrate how Linked Data facilitates metadata interoperability and cross-institutional resource sharing (Haslhofer et al., 2019).

The integration of Knowledge Graphs with library catalogs has significantly improved search functionalities and subject classification. By leveraging machine-learning-enhanced entity recognition, libraries can automate metadata enrichment and enhance knowledge discovery through linked concepts and relationships (Lüschow, 2022). For example, ontology-driven knowledge graphs have been used to visualize research domains in Chinese academic libraries, facilitating more structured resource categorization (Lu & Jimei, 2024).

However, the implementation of Knowledge Graphs in libraries is not without challenges. Data quality issues, inconsistent metadata standards, and technical integration complexities pose obstacles to seamless adoption (Haller et al., 2020). Moreover, maintaining dynamic linkages between evolving datasets requires continuous updates and robust entity reconciliation mechanisms (Jia, 2020).

### 2.6 Linked Data in Open Science and FAIR Principles

Open Science promotes transparency, accessibility, and reproducibility in research, with Linked Data serving as a fundamental component in enabling interoperability and data sharing across disciplines (Umbach, 2024). The principles of Open Science align with the use of Linked Data, as it allows for semantic structuring and linking of research datasets, making them discoverable and reusable in multiple contexts (Mons et al., 2017).

Linked Data enhances cross-disciplinary data integration, addressing the challenge of siloed data repositories in research institutions. By using RDF (Resource Description Framework) and SPARQL (a query language for Linked Data), research data can be structured and linked to external knowledge graphs, ensuring semantic interoperability (Larsson et al., 2025). This approach has been instrumental in biomedical research, climate science, and digital humanities, where large datasets require structured integration for better knowledge discovery (Hasnain & Rebholz-Schuhmann, 2018).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

The FAIR (Findable, Accessible, Interoperable, Reusable) Principles were introduced to enhance the usability and accessibility of research data (Wilkinson et al., 2016). Linked Data directly supports these principles by:

- Findability: Assigning persistent identifiers (e.g., DOIs, URIs) to research datasets, making them machine-readable and easily searchable.
- Accessibility: Ensuring open access to data through standardized Linked Data protocols, facilitating seamless data retrieval (Dunning et al., 2017).
- Interoperability: Using common metadata standards (e.g., Dublin Core, Schema.org, CIDOC CRM) to allow integration across different research domains (Jacob et al., 2025).
- Reusability: Enhancing dataset documentation and semantic annotation, making data reusable for new research purposes (Bhat & Wani, 2025).

The European Open Science Cloud (EOSC) has been instrumental in advancing FAIR-compliant Linked Data infrastructures. EOSC fosters cross-border data sharing, leveraging Linked Data technologies to aggregate datasets from various European research institutions (Mons et al., 2017).

### 2.7 Implementation Challenges and Solutions

Despite its advantages, integrating Linked Data with Open Science and FAIR principles faces technical and organizational challenges. These include:

- Heterogeneous Data Standards: Different scientific disciplines use diverse metadata schemas, complicating interoperability. Schema alignment frameworks, such as FAIR Digital Objects (FDOs), aim to standardize metadata representation (Soiland-Reyes et al., 2024).
- Scalability Issues: Managing large-scale Linked Data infrastructures requires high-performance computing and cloud-based storage solutions (Wilkinson et al., 2025).
- Data Quality and Provenance: Ensuring the accuracy, completeness, and provenance tracking of Linked Data remains a critical challenge (Hasnain & Rebholz-Schuhmann, 2018). Machine-learning-driven metadata validation tools are being developed to enhance data integrity and consistency (Wilkinson et al., 2016).

### 2.8 Scientometric Insights into Linked Data Research

The evolution of Linked Data (LD) as a framework for publishing and interconnecting structured data on the Web has significantly influenced research and practice across various disciplines, particularly in information science, library services, and digital governance. Initiated by Tim Berners-Lee’s foundational principles in 2006, LD emphasizes using Uniform Resource Identifiers (URIs), Resource Description Framework (RDF), and SPARQL to create a decentralized and interconnected web of data. Over the past decade, the academic community has shown substantial interest in LD, resulting in a sharp rise in publications, diverse thematic developments, and the emergence of complex applications (Gupta et al. 2020).

Bibliometric analyses illustrate the multidimensional nature of LD research, with the semantic web and ontology consistently appearing as dominant clusters in co-word and thematic mapping. In this regard, Hosseini et al. (2025) identify nine major co-occurrence topic clusters and twenty-nine thematic clusters, emphasizing mature themes like ontology-based deep learning, semantic bioinformatics, and knowledge representation.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

These developments underline LD's foundational role in structuring and discovering complex knowledge systems. Similarly, Wu and Ye (2021) observe that semantic web and gene ontology serve as two of the most influential areas, with interdisciplinary applications ranging from natural language processing to bioinformatics.

The application of LD in libraries represents one of its most tangible institutional adoptions. Gaitanou et al. (2022) argue that libraries are transitioning from isolated data silos to open, interoperable systems by leveraging LD technologies to publish metadata in machine-readable formats. Projects such as BIBFRAME and the use of RDF vocabularies allow for enhanced discoverability of library resources across global networks.

However, this transition is not without its difficulties. Wahi et al. (2018) critically highlight the persistence of outdated metadata standards like MARC, technical complexity of LD tools, and inconsistencies in controlled vocabularies as major barriers to full-scale adoption. These limitations suggest that while LD presents opportunities for systemic innovation, infrastructural inertia and lack of training hinder its effective deployment.

In the domain of health information management, LD has proven essential for addressing the fragmentation of healthcare data. Costa et al. (2024) provide a bibliometric overview indicating that health data systems suffer from a lack of interoperability and poor linkage across platforms. Through the lens of LD, the integration of disparate datasets—ranging from patient records to death certificates—becomes achievable, supporting more comprehensive and efficient information governance. Nevertheless, their findings also point to a residual gap in the correlation between interoperability studies and practical linkage applications, underscoring the need for more cohesive frameworks and empirical models.

In scientometric and e-science contexts, LD's impact is both functional and methodological. Narock and Wimmer (2017) assess the adoption of semantic technologies in geosciences and observe that while the field has embraced semantic e-science frameworks, there is often a disconnect between theoretical Semantic Web advancements and their application in domain-specific problems. This challenge is reflective of a broader issue in the LD landscape: the lag between technological innovation and its contextual integration within specialized disciplines.

Open data movements also benefit from LD through enhanced transparency and accessibility. Zhang et al. (2017) trace the trajectory of open data research, noting emerging themes such as crowdsourcing, data journalism, and big data governance. LD not only structures these data but enables linking across projects and institutions, thereby fostering collaborative innovation. Similarly, Lytras, Hassan, and Aljohani (2019) emphasize the role of LD in smart library services, where personalized content delivery and social network analytics transform user engagement in urban and educational environments.

Despite its advantages, LD's ecosystem remains fragmented. Studies consistently point to issues such as lack of universal standards, semantic ambiguity, and steep learning curves for tool adoption. Moreover, the theoretical richness of LD is not always matched by practical utility, a point reiterated across multiple analyses (Hosseini et al., 2021). This critical gap demands more user-centered research, training, and tool development to bridge the divide between concept and application.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

# 3. METHODOLOGY

## 3.1 Research Design

This study adopts a quantitative bibliometric research design to systematically investigate the structure, evolution, and scholarly impact of Linked Data (LD) research from 2005 to 2024. Bibliometric analysis allows for the objective measurement of scientific productivity and influence by analyzing publication outputs, citation patterns, collaboration networks, and thematic trends. The study applies both performance analysis (e.g., citation counts, publication volume) and science mapping techniques (e.g., co-authorship analysis, keyword co-occurrence, thematic clustering) to generate insights into the intellectual development of the LD field. This dual approach ensures a comprehensive assessment of both the historical foundations and emerging directions in LD research.

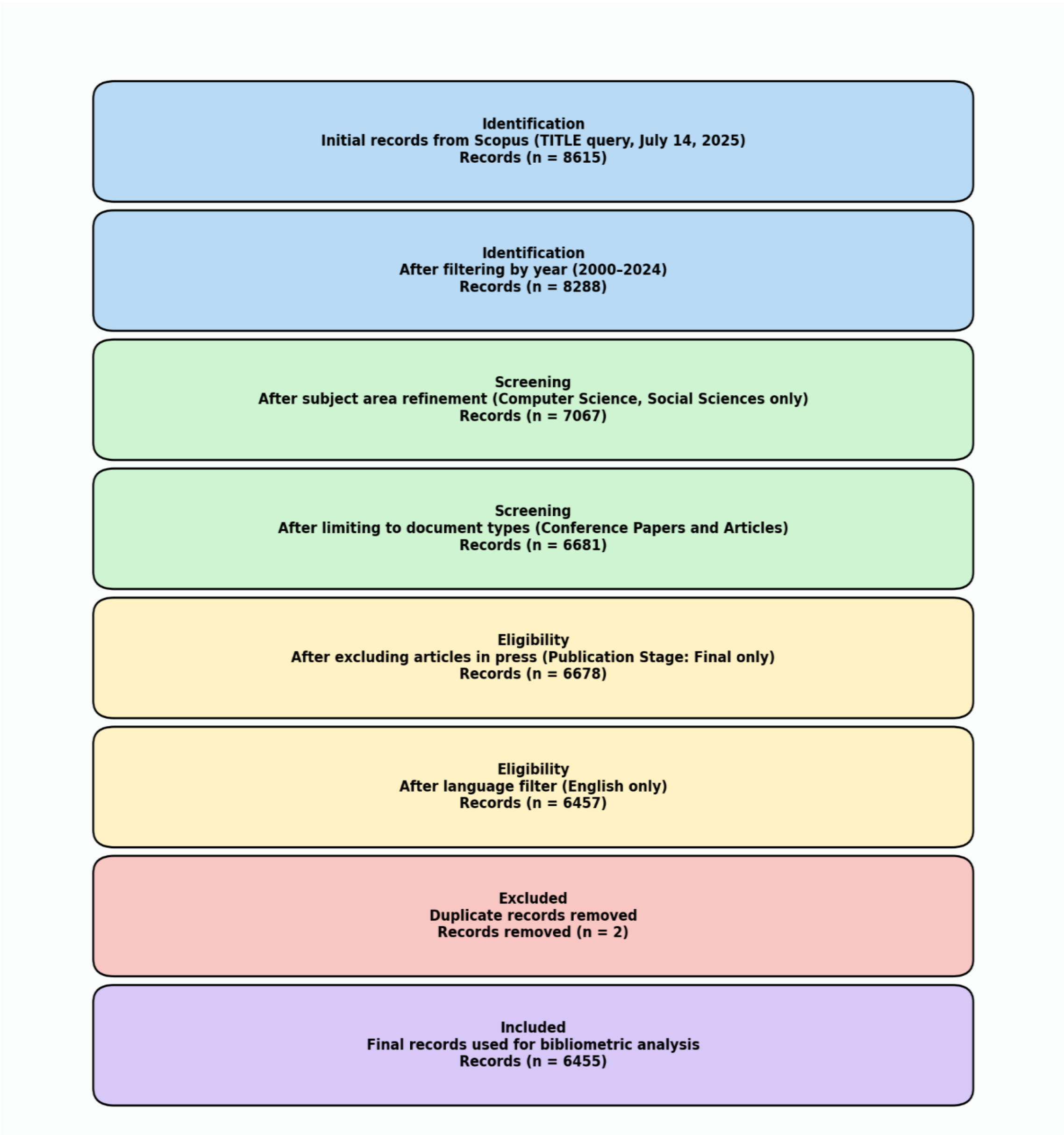


**Figure 1: PRISMA Flow Diagram Representing the Bibliometric Record Selection Process**

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

The dataset used for this study was retrieved from the Scopus database on July 14, 2025. A carefully constructed TITLE-based search query was employed to identify publications that explicitly addressed Linked Data concepts, technologies, and applications. The search query included a combination of terms such as "Linked Data," "RDF," "SPARQL," "Semantic Web," "FAIR Data," "Knowledge Graphs," "Metadata Interoperability," and "Open Data," among others. To ensure a focused and relevant dataset, the search was limited to documents published between 2000 and 2024, with only two document types selected: journal articles and conference papers. The query was restricted to final-stage publications written in English.

Importantly, the subject areas were filtered to include only Computer Science and Social Sciences. Computer Science was selected due to its foundational role in developing the technical underpinnings of Linked Data technologies, such as RDF (Resource Description Framework), SPARQL, and graph databases. Social Sciences were included because many of the real-world applications of LD—such as digital governance, open knowledge systems, digital inclusion, and information behavior—are addressed within this domain. Figure 1 outlines the full set of inclusion and exclusion criteria used in the study, furthermore the intersection of these two disciplines reflects the multidisciplinary nature of LD research, which spans both the technological development and its social or institutional implementation.

Following a structured PRISMA-based screening process, the initial set of 8,615 documents was refined through a series of filters: limiting the publication years, document types, subject areas, publication stages, and language. After removing duplicates and ineligible entries, the final dataset consisted of 6,455 documents, forming the empirical basis of the bibliometric analysis.

### 3.3 Data Analysis Methods

The analysis was carried out using a combination of open-source and specialized bibliometric tools, including the Bibliometrix R package, VOSviewer, and Microsoft Excel. Bibliometrix was employed for data preprocessing, statistical analysis, and science mapping, while VOSviewer was used to generate visualizations such as co-authorship networks, keyword co-occurrence maps, and thematic clusters. Descriptive statistics were first used to assess overall productivity trends, including annual growth rates, total publications, and citation impact metrics such as average citations per document. Author-level productivity was evaluated using metrics like the h-index, g-index, and m-index. Source impact was analyzed through Bradford's Law, while highly cited documents were examined through both global and local citation metrics.

To understand the thematic evolution of the field, keyword co-occurrence analysis was performed using both author keywords and Keywords Plus. These were further grouped into clusters to reveal dominant themes and conceptual overlaps. Thematic mapping was carried out using Callon's centrality and density measures to classify themes into basic, motor, niche, and emerging or declining categories. Bibliographic coupling and factorial analysis were used to uncover intellectual relationships among highly cited documents and to map the structure of conceptual clusters. Co-authorship networks were also visualized at both author and country levels to assess collaboration patterns and identify influential nodes in the global research network.

### 3.4 Ethical Considerations

This study is entirely based on secondary data retrieved from publicly accessible bibliographic databases and does not involve any interaction with human subjects. All

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

bibliographic metadata (e.g., author names, titles, journal information, citation data) were used in accordance with academic fair use policies. No personally identifiable or sensitive information was collected or processed during the study. The research adheres to ethical standards of scientific reporting, transparency, and attribution, and all cited sources have been properly acknowledged.

### 3.5 Study Limitations

Despite its comprehensive approach, this study is subject to certain limitations. First, it relies solely on the Scopus database, which, while extensive, may not cover all relevant publications indexed in other repositories like Web of Science, IEEE Xplore, or Google Scholar. This might result in the omission of some impactful but unindexed documents. Second, the TITLE-based search strategy, though precise, may have excluded papers that discuss Linked Data extensively but do not mention key terms in the title. Third, bibliometric indicators such as citation counts can be biased by self-citations, field-specific citation behavior, or publication age. Finally, while the study includes Computer Science and Social Sciences to reflect the interdisciplinary nature of LD research, other potentially relevant domains like Health Informatics, Education, or Environmental Studies were not included due to scope constraints.

### 3.4 Ethical Considerations

This study is based on publicly available bibliographic data and does not involve human subjects. All citations and sources used in this study are properly attributed, ensuring academic integrity and compliance with ethical research standards. No personally identifiable information is included, and the analysis adheres to best practices in bibliometric research.

## 4. RESULTS

**Table 1: Descriptive Summary of the Bibliometric Dataset on Linked Data Research (2000–2024)**

| Description | Results |
|---|---|
| Timespan | 2000-2024 |
| Sources (Journals, Books, etc) | 1479 |
| Documents | 6455 |
| Annual Growth Rate % | 13.22 |
| Document Average Age | 9.81 |
| Average citations per doc | 12.66 |
| References | 115230 |
| **Document Contents** | |
| Keywords Plus (ID) | 16735 |
| Author's Keywords (DE) | 8043 |
| **Authors** | |
| Authors | 10211 |
| Authors of single-authored docs | 547 |
| **Authors Collaboration** | |
| Single-authored docs | 662 |
| Co-Authors per Doc | 3.43 |
| International co-authorships % | 22.35 |
| **Document Types** | |
| Article | 1542 |
| Conference paper | 4913 |

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

Table 1 presents a comprehensive overview of the bibliometric dataset used in this study, covering the timespan from 2000 to 2024. A total of 6,455 documents were analyzed, drawn from 1,479 unique sources including journals, conference proceedings, and books. The data reflect a healthy annual growth rate of 13.22%, indicating sustained scholarly interest in linked data research over the 25-year period.

The average document age is approximately 9.81 years, and each document received on average 12.66 citations, highlighting a moderate but consistent impact. The reference pool is extensive, totaling over 115,000 references, which suggests a well-established and interconnected body of literature.

In terms of content, the dataset includes 16,735 Keywords Plus (index terms) and 8,043 author-supplied keywords, indicating rich thematic diversity. Authorship analysis reveals contributions from 10,211 unique authors, of which 547 authored single-authored documents, while the majority of work was collaborative. On average, each paper had 3.43 co-authors, and 22.35% of the publications involved international collaboration, underscoring the global nature of the research community.

Regarding document types, the majority are conference papers (4,913), with articles (1,542) forming the remainder, highlighting the importance of conference venues in disseminating research on linked data and semantic web technologies.

**Table 2: Annual Scientific Production in Linked Data Research (2000–2024)**

| Year | Articles | Year | Articles | Year | Articles |
|---|---|---|---|---|---|
| 2000 | 9 | 2010 | 272 | 2020 | 308 |
| 2001 | 15 | 2011 | 282 | 2021 | 265 |
| 2002 | 29 | 2012 | 393 | 2022 | 221 |
| 2003 | 34 | 2013 | 537 | 2023 | 196 |
| 2004 | 32 | 2014 | 580 | 2024 | 177 |
| 2005 | 83 | 2015 | 606 | | |
| 2006 | 92 | 2016 | 570 | | |
| 2007 | 89 | 2017 | 517 | | |
| 2008 | 138 | 2018 | 452 | | |
| 2009 | 169 | 2019 | 389 | | |

Table 2 outlines the annual scientific output in linked data research from 2000 to 2024, illustrating the field's steady growth and eventual stabilization. The early years (2000–2004) show a modest start, with fewer than 35 publications per year, indicating the nascent stage of research in this domain. Beginning in 2005, there is a notable upward trend, with publication counts increasing significantly, peaking in 2015 with 606 articles—the highest in the 25-year span.

From 2010 onward, the field experiences a surge in scholarly output, consistently producing over 250 articles per year. This period reflects heightened academic and practical interest in linked data, semantic web technologies, and their applications. While the number of publications gradually declined after the 2015 peak, annual production remained robust, with over 200 publications per year through 2022. The slight decrease observed in 2023 (196 articles) and 2024 (177 articles) may reflect the natural saturation or shifting research priorities.

This table evidences the maturity and sustained relevance of linked data research, particularly in the last decade, and reflects its establishment as a significant area within information and computer sciences.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

**Table 3: Citation Impact Metrics by Year for Linked Data Publications (2000–2024)**

| Year | MeanTCperArt | N | MeanTCperYear | CitableYears |
|---|---|---|---|---|
| 2000 | 80.67 | 9 | 3.1 | 26 |
| 2001 | 50.6 | 15 | 2.02 | 25 |
| 2002 | 78.48 | 29 | 3.27 | 24 |
| 2003 | 64.24 | 34 | 2.79 | 23 |
| 2004 | 20.72 | 32 | 0.94 | 22 |
| 2005 | 17.16 | 83 | 0.82 | 21 |
| 2006 | 14.5 | 92 | 0.72 | 20 |
| 2007 | 22.11 | 89 | 1.16 | 19 |
| 2008 | 22.53 | 138 | 1.25 | 18 |
| 2009 | 44.31 | 169 | 2.61 | 17 |
| 2010 | 17.17 | 272 | 1.07 | 16 |
| 2011 | 20.33 | 282 | 1.36 | 15 |
| 2012 | 17.53 | 393 | 1.25 | 14 |
| 2013 | 14.64 | 537 | 1.13 | 13 |
| 2014 | 12.99 | 580 | 1.08 | 12 |
| 2015 | 8.33 | 606 | 0.76 | 11 |
| 2016 | 10.64 | 570 | 1.06 | 10 |
| 2017 | 8.68 | 517 | 0.96 | 9 |
| 2018 | 8.12 | 452 | 1.01 | 8 |
| 2019 | 7.24 | 389 | 1.03 | 7 |
| 2020 | 7.58 | 308 | 1.26 | 6 |
| 2021 | 4.15 | 265 | 0.83 | 5 |
| 2022 | 4.14 | 221 | 1.03 | 4 |
| 2023 | 2.67 | 196 | 0.89 | 3 |
| 2024 | 0.75 | 177 | 0.38 | 2 |

Table 3 provides an in-depth view of the citation impact of publications related to linked data over a 25-year period (2000–2024). The column Mean TC per Article indicates the average total citations per article for each publication year, while Mean TC per Year reflects the average yearly citation rate, normalized by the number of years each article has been citable.

The data clearly show that earlier publications (2000–2003) have significantly higher citation averages, with 2000 (80.67) and 2002 (78.48) leading, indicating the long-term influence of foundational works. As expected, Mean TC per Article tends to decline over time, due to fewer years of citation accumulation for more recent publications.

In terms of Mean TC per Year, earlier years such as 2002 (3.27) and 2003 (2.79) again stand out, demonstrating strong and sustained citation activity. After 2015, both citation metrics begin to decline gradually, with the most recent years (especially 2023–2024) showing lower impact, which is typical due to the short citation window.

This table highlights the long-lasting citation influence of early research in the domain and emphasizes the need for time to fully evaluate the impact of newer studies. It also underscores how early foundational works continue to shape the intellectual structure of the field.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

**Table 4: Most Relevant Sources in Linked Data Research Publications (2000–2024)**

| Sources | Articles |
|---|---|
| Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 1278 |
| CEUR Workshop Proceedings | 1231 |
| ACM International Conference Proceeding Series | 256 |
| Communications in Computer and Information Science | 190 |
| Semantic Web | 89 |
| Proceedings of the International Conference on Dublin Core and Metadata Applications | 56 |
| Journal of Web Semantics | 49 |
| Lecture Notes in Business Information Processing | 37 |
| Advances in Intelligent Systems and Computing | 35 |
| International Journal on Semantic Web and Information Systems | 35 |

Table 4 lists the most prolific sources contributing to the body of literature on linked data research between 2000 and 2024. The Lecture Notes in Computer Science (LNCS) series leads significantly with 1,278 publications, reflecting its strong association with conference-based scholarly communication in computer science and related disciplines. Closely following is the CEUR Workshop Proceedings with 1,231 articles, indicating the importance of workshops and informal academic gatherings in shaping the discourse around linked data. Other major contributors include the ACM International Conference Proceeding Series (256 articles) and Communications in Computer and Information Science (190 articles), both of which also focus on conference outputs, reinforcing the field's conference-driven publication culture.

Notably, peer-reviewed journals like Semantic Web (89 articles) and the Journal of Web Semantics (49 articles) provide platforms for in-depth and high-quality research, albeit in smaller quantities compared to proceedings. The International Journal on Semantic Web and Information Systems, while smaller in count (35 articles), represents a specialized venue aligned closely with the topic. This table highlights the prominent role of conference proceedings in disseminating linked data research and the key outlets where scholars in the field regularly publish their work.

**Table 5: Core Sources in Linked Data Research Identified by Bradford's Law (2000–2024)**

| SO | Rank | Freq | cumFreq | Zone |
|---|---|---|---|---|
| Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 1 | 1278 | 1278 | Zone 1 |
| CEUR Workshop Proceedings | 2 | 1231 | 2509 | Zone 1 |
| ACM International Conference Proceeding Series | 3 | 256 | 2765 | Zone 2 |
| Communications in Computer and Information Science | 4 | 190 | 2955 | Zone 2 |
| Semantic Web | 5 | 89 | 3044 | Zone 2 |
| Proceedings of the International Conference on Dublin Core and Metadata Applications | 6 | 56 | 3100 | Zone 2 |
| Journal of Web Semantics | 7 | 49 | 3149 | Zone 2 |
| Lecture Notes in Business Information Processing | 8 | 37 | 3186 | Zone 2 |
| Advances in Intelligent Systems and Computing | 9 | 35 | 3221 | Zone 2 |
| International Journal on Semantic Web and Information Systems | 10 | 35 | 3256 | Zone 2 |

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

Table 5 presents the core sources of linked data research as identified using Bradford's Law of Scattering, which categorizes sources into productivity zones based on publication frequency. According to this law, a small number of journals (Zone 1) contribute a disproportionately large number of articles, while subsequent zones include progressively more journals with fewer articles.

The analysis reveals that two sources dominate Zone 1:

- Lecture Notes in Computer Science (1,278 articles)
- CEUR Workshop Proceedings (1,231 articles)

Together, these two outlets account for 2,509 articles, emphasizing their central role in publishing linked data research.

Zone 2 comprises eight sources, each contributing fewer articles but still collectively significant. These include ACM International Conference Proceeding Series, Communications in Computer and Information Science, and Semantic Web, among others. The cumulative total in Zone 2 reaches 3,256 articles, demonstrating the broader distribution of scholarly output across secondary sources.

This distribution aligns with Bradford's Law, affirming the high concentration of literature in a few core outlets, while indicating the presence of a supporting layer of journals and conferences that sustain the field's breadth and diversity. It highlights the conference-centric nature of the linked data research community, particularly in computer science and information systems.

**Table 6: Local Impact of Sources in Linked Data Research: Citation and Index Analysis (2000–2024)**

| Element | h_index | g_index | m_index | TC | NP | PY_start |
|---|---|---|---|---|---|---|
| Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) | 59 | 100 | 2.269 | 19934 | 1278 | 2000 |
| CEUR Workshop Proceedings | 26 | 40 | 1.13 | 4555 | 1231 | 2003 |
| Semantic Web | 24 | 47 | 1.714 | 2467 | 89 | 2012 |
| Journal of Web Semantics | 23 | 45 | 1.353 | 2085 | 49 | 2009 |
| ACM International Conference Proceeding Series | 21 | 38 | 1.05 | 2351 | 256 | 2006 |
| Proceedings of the VLDB Endowment | 17 | 25 | 0.944 | 2261 | 25 | 2008 |
| Communications in Computer and Information Science | 14 | 24 | 0.875 | 1047 | 190 | 2010 |
| Proceedings – International Conference on Data Engineering | 13 | 30 | 0.619 | 958 | 34 | 2005 |
| IEEE Transactions on Knowledge and Data Engineering | 12 | 16 | 0.522 | 734 | 16 | 2003 |
| International Journal on Semantic Web and Information Systems | 12 | 35 | 0.6 | 3922 | 35 | 2006 |

Table 6 evaluates the local scholarly impact of leading sources in linked data research through key citation-based metrics: h-index, g-index, m-index, total citations (TC), number of publications (NP), and the publication start year (PY_start). These indicators reflect not only publication volume but also influence and consistency over time.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

The Lecture Notes in Computer Science (LNCS) stands out with the highest h-index (59) and g-index (100), as well as the highest total citations (19,934), emphasizing its dominant role in the field. Its m-index (2.269)—which normalizes the h-index by the number of active publication years—further confirms LNCS's sustained scholarly impact since 2000.

Other notable contributors include the CEUR Workshop Proceedings, which, despite a lower citation count (4,555), maintains a strong publication volume (1,231) and a respectable h-index (26), reflecting its role as a key outlet for emerging research. Journals such as Semantic Web, Journal of Web Semantics, and the International Journal on Semantic Web and Information Systems show strong m-indices (≥1), suggesting high-quality outputs with relatively fewer publications.

Interestingly, the International Journal on Semantic Web and Information Systems, with only 35 publications, has accrued nearly 4,000 citations, highlighting its impact despite modest productivity. Similarly, IEEE Transactions on Knowledge and Data Engineering shows high citation-per-paper value, reflecting its prestige and selective publication strategy.

This table illustrates the dual nature of impact in linked data research: one driven by volume and visibility in conference-oriented outlets, and another by selective, high-impact journal publishing.

**Table 7: Most Relevant Authors in Linked Data Research (2000–2024)**

| Authors | Articles | Articles Fractionalized |
|---|---|---|
| Auer, S. | 73 | 18.06 |
| Verborgh, R. | 68 | 15.17 |
| Hyvönen, E. | 57 | 16.35 |
| Mannens, E. | 55 | 10.08 |
| Lehmann, J. | 51 | 11.63 |
| Ngomo, A.-C. N. | 49 | 11.83 |
| Wang, X. | 48 | 12.11 |
| Polleres, A. | 43 | 10.57 |
| Harth, A. | 41 | 13.37 |
| Decker, S. | 38 | 9.27 |

Table 7 presents the most prolific authors in the field of linked data research from 2000 to 2024, based on two metrics: the total number of articles and the fractionalized count—which accounts for co-authorship by dividing credit proportionally among all authors of each paper.

Sören Auer leads the field with 73 publications and a substantial fractionalized contribution of 18.06, indicating both high productivity and frequent collaboration. Ruben Verborgh and Eero Hyvönen follow closely with 68 and 57 publications respectively, with Hyvönen showing a relatively high fractionalized contribution (16.35), suggesting more lead or solo-authored works.

Other key contributors include Mannens, Lehmann, Ngomo, Wang, Polleres, Harth, and Decker, each having published between 38 and 55 articles. Notably, Harth, A. stands out with a strong fractionalized score (13.37) relative to his 41 publications, implying a significant authorial role in fewer, more focused contributions.

These authors collectively represent a core intellectual network within the linked data research community, often associated with leading projects, institutions, and standards

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

development. Their consistent contributions underscore their influence in shaping the theoretical and practical foundations of the field.

**Table 8: Local Impact of Authors in Linked Data Research: Citation and Index Metrics (2000–2024)**

| Element | h_index | g_index | m_index | TC | NP | PY_start |
|---|---|---|---|---|---|---|
| Auer, S. | 23 | 48 | 1.15 | 2392 | 73 | 2006 |
| Polleres, A. | 20 | 40 | 1.053 | 1666 | 43 | 2007 |
| Lehmann, J. | 19 | 49 | 1.118 | 2449 | 51 | 2009 |
| Bizer, C. | 17 | 28 | 0.81 | 6716 | 28 | 2005 |
| Decker, S. | 17 | 38 | 0.654 | 2087 | 38 | 2000 |
| Hogan, A. | 16 | 31 | 1 | 1076 | 31 | 2010 |
| Paulheim, H. | 16 | 30 | 1.067 | 1529 | 30 | 2011 |
| Harth, A. | 15 | 33 | 0.714 | 1126 | 41 | 2005 |
| Ngomo, A.-C. N. | 15 | 33 | 1.071 | 1164 | 49 | 2012 |
| Cimiano, P. | 14 | 26 | 0.824 | 1246 | 26 | 2009 |

Table 8 presents a comparative analysis of the local impact of key authors in the field of linked data research based on citation and productivity indices. The h-index reflects both productivity and citation impact, while the g-index gives more weight to highly cited publications. The m-index, which normalizes the h-index by the number of years since the first publication, helps highlight consistent academic contribution over time.

Auer, S. emerges as a leading figure, with the highest publication count (73), a strong h-index of 23, and the highest m-index (1.15) among early-career contributors, indicating sustained influence and productivity since 2006. Lehmann, J. also shows a notable g-index of 49 and a high citation total (2,449), underscoring his impact in a relatively short period.

Interestingly, Bizer, C., despite having only 28 publications, leads in total citations (6,716), reflecting his foundational contributions and the enduring relevance of a smaller number of highly influential works.

Authors like Polleres, Harth, and Decker show long-standing engagement with moderate but steady impact, while more recent contributors such as Ngomo, Paulheim, and Hogan exhibit strong m-indices (≥1), signaling rapidly growing influence.

This table highlights a blend of foundational scholars with highly cited seminal works and emerging leaders with consistent high-impact outputs, painting a diverse and dynamic portrait of scholarly leadership in the linked data research community.

**Table 9: Corresponding Author's Countries in Linked Data Research: Scientific Contribution and Collaboration Patterns (2000–2024)**

| Country | Articles | SCP | MCP | Freq | MCP_Ratio |
|---|---|---|---|---|---|
| China | 361 | 5.6 | 315 | 46 | 12.7 |
| Germany | 360 | 5.6 | 299 | 61 | 16.9 |
| USA | 354 | 5.5 | 313 | 41 | 11.6 |
| Italy | 210 | 3.3 | 160 | 50 | 23.8 |
| France | 196 | 3 | 156 | 40 | 20.4 |
| United Kingdom | 161 | 2.5 | 122 | 39 | 24.2 |
| Spain | 137 | 2.1 | 92 | 45 | 32.8 |
| Greece | 131 | 2 | 100 | 31 | 23.7 |
| Japan | 99 | 1.5 | 90 | 9 | 9.1 |
| Korea | 98 | 1.5 | 88 | 10 | 10.2 |

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

Table 9 summarizes the scientific contribution and international collaboration patterns in linked data research based on the country affiliation of corresponding authors. The data reveals both the volume of scholarly output and the extent of global collaboration.

China, Germany, and the USA lead in terms of total articles, each with over 350 publications, contributing approximately 5.5%–5.6% of the total output. These countries also show strong international engagement, especially Germany with the highest number of multiple country publications (MCP = 299) and an MCP ratio of 16.9%, indicating a robust collaborative research environment.

Italy, France, the UK, Spain, and Greece follow as mid-level contributors. While their total output is lower, they demonstrate significantly higher MCP ratios, particularly Spain (32.8%) and the United Kingdom (24.2%), suggesting a strong inclination toward cross-border collaborations.

Japan and Korea, although producing a moderate number of articles (99 and 98, respectively), have the lowest MCP ratios (9.1% and 10.2%), reflecting more domestically focused research efforts.

This table illustrates that while scientific volume is concentrated in a few large countries, international collaboration is more evenly distributed, with several European countries showing leadership in building global research networks within the linked data domain.

**Table 10: Most Cited Countries in Linked Data Research: Total Citations and Average Citations per Article (2000–2024)**

| Country | TC | Average Article Citations |
|---|---|---|
| Germany | 10493 | 29.1 |
| USA | 5884 | 16.6 |
| United Kingdom | 4368 | 27.1 |
| Netherlands | 3222 | 33.2 |
| Italy | 2725 | 13 |
| Ireland | 2614 | 27.5 |
| China | 2524 | 7 |
| Greece | 1984 | 15.1 |
| Spain | 1807 | 13.2 |
| France | 1721 | 8.8 |

Table 10 highlights the top contributing countries in terms of citational impact in linked data research from 2000 to 2024, focusing on total citation count and average citations per article.

Germany leads with 10,493 total citations and a strong average of 29.1 citations per article, confirming its foundational and influential role in shaping the field. The United States, while second in total citations (5,884), shows a relatively lower average (16.6 citations per article), likely reflecting its broader research volume but slightly more dispersed impact.

The Netherlands achieves the highest average citations per article (33.2), despite a smaller publication count, indicating high-impact contributions. Similarly, Ireland and the United Kingdom also demonstrate strong averages (27.5 and 27.1 respectively), suggesting that research from these countries is widely referenced and influential.

On the other hand, countries like China, France, and Italy, though among the top in output, have lower average citations (7.0, 8.8, and 13.0 respectively), pointing to potential gaps in visibility, language accessibility, or impact orientation of their publications.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

This analysis underlines that while volume of publications is important, citation averages offer critical insight into scholarly influence and recognition, with countries like the Netherlands and Ireland exemplifying quality over quantity in linked data research.

**Table 11: Most Globally Cited Documents in Linked Data Research: Citation Impact and Influence (2000–2024)**

| Paper | DOI | Total Citations | TC per Year | Normalized TC |
|---|---|---|---|---|
| Bizer, C. (2009). International Journal on Semantic Web and Information Systems | 10.4018/jswis.2009081901 | 3411 | 200.65 | 76.97 |
| Heath, T. (2011). Synthesis Lectures on the Semantic Web: Theory and Technology | 10.2200/S00334ED1V01Y201102WBE001 | 1199 | 79.93 | 58.98 |
| Horrocks, I. (2003). Web Semantics | 10.1016/j.websem.2003.07.001 | 1161 | 50.48 | 18.07 |
| Broekstra, J. (2002). Lecture Notes in Computer Science | 10.1007/3-540-48005-6_7 | 841 | 35.04 | 10.72 |
| Nejdl, W. (2002). Proceedings of the International Conference on World Wide Web (WWW) | 10.1145/511446.511525 | 516 | 21.5 | 6.57 |
| Zaveri, A. (2016). Semantic Web | 10.3233/SW-150175 | 480 | 48 | 45.13 |
| Decker, S. (2000). IEEE Internet Computing | 10.1109/4236.877487 | 464 | 17.85 | 5.75 |
| Neumann, T. (2010). VLDB Journal | 10.1007/s00778-009-0165-y | 462 | 28.88 | 26.91 |
| Neumann, T. (2008). Proceedings of the VLDB Endowment | 10.14778/1453856.1453927 | 461 | 25.61 | 20.46 |
| Unger, C. (2012). Proceedings of the Annual Conference on World Wide Web (WWW) | 10.1145/2187836.2187923 | 388 | 27.71 | 22.13 |

Table 11 lists the most globally cited documents in linked data research, ranking them by total citation count, annual citation rate, and normalized citations, which adjust for field and publication year.

The most influential publication by far is Bizer (2009), with 3,411 total citations and a remarkable 200.65 citations per year, making it a foundational work in the domain. Following it, Heath (2011) and Horrocks (2003) also stand out with over 1,100 citations each, contributing substantially to theoretical and technical frameworks in the field.

Zaveri (2016), though more recent, demonstrates strong influence with a normalized TC of 45.13, indicating rapid uptake and relevance. Similarly, Neumann's works (2008 and 2010) and Unger (2012) show sustained impact over time, reflected in both annual and normalized citation scores.

These highly cited documents reflect the intellectual backbone of the linked data field, encompassing early architectural proposals, theoretical advancements, and evaluative frameworks that continue to shape research trajectories.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

**Table 12: Most Locally Cited Documents in Linked Data Research: Citation Impact and Influence (2000–2024)**

| Document | DOI | Year | Local Citations | Global Citations | LC/GC Ratio (%) | Normalized Local Citations | Normalized Global Citations |
|---|---|---|---|---|---|---|---|
| Bizer, C. (2009). International Journal on Semantic Web and Information Systems | 10.4018/jswis.2009081901 | 2009 | 680 | 3411 | 19.94 | 80.31 | 76.97 |
| Neumann, T. (2010). VLDB Journal | 10.1007/s00778-009-0165-y | 2010 | 163 | 462 | 35.28 | 51.37 | 26.91 |
| Neumann, T. (2008). Proceedings of the VLDB Endowment | 10.14778/1453856.1453927 | 2008 | 149 | 461 | 32.32 | 31.73 | 20.46 |
| Broekstra, J. (2002). Lecture Notes in Computer Science | 10.1007/3-540-48005-6_7 | 2002 | 126 | 841 | 14.98 | 16.17 | 10.72 |
| Huang, J. (2011). Proceedings of the VLDB Endowment | 10.14778/3402707.3402747 | 2011 | 112 | 371 | 30.19 | 35.29 | 18.25 |
| Quilitz, B. (2008). Lecture Notes in Computer Science | 10.1007/978-3-540-68234-9_39 | 2008 | 91 | 343 | 26.53 | 19.38 | 15.22 |
| Schmachtenberg, M. (2014). Lecture Notes in Computer Science | 10.1007/978-3-319-11964-9 | 2014 | 90 | 282 | 31.91 | 49.06 | 21.71 |
| Zeng, K. (2013). Proceedings of the VLDB Endowment | 10.14778/2535570.2488333 | 2013 | 88 | 252 | 34.92 | 39.25 | 17.22 |
| Le-Phuoc, D. (2011). Lecture Notes in Computer Science | 10.1007/978-3-642-25073-6_24 | 2011 | 80 | 333 | 24.02 | 25.21 | 16.38 |
| Hartig, O. (2009). Lecture Notes in Computer Science-A | 10.1007/978-3-642-04930-9_19 | 2009 | 73 | 207 | 35.27 | 8.62 | 4.67 |

Table 12 showcases the most locally cited documents in the field of linked data research, where “local citations” refer to how frequently a document is cited within the set of 6,457 documents used in this bibliometric analysis. This local impact is compared against global citation counts and normalized citation indicators.

Bizer (2009) remains the most influential both globally and locally, with 680 local citations out of 3,411 global citations, underscoring its foundational role and sustained

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

relevance within the linked data community. Its normalized local citation score (80.31) is the highest, confirming deep engagement within the field.

Papers by Neumann (2008, 2010) also show strong dual impact, with local-to-global citation ratios exceeding 30%, indicating their centrality to linked data research discussions. Similarly, works by Huang, Schmachtenberg, and Zeng also demonstrate high LC/GC ratios, suggesting focused influence within the core research domain.

Interestingly, while some papers like Hartig (2009) have moderate global reach (207 citations), their LC/GC ratio is high (35.27%), meaning they are especially foundational for subsequent studies within the same community, even if less cited externally.

This table highlights documents that are not only widely recognized globally but also deeply embedded in the thematic evolution and methodological backbone of the linked data research ecosystem. These works serve as pillars for internal scholarly dialogue and academic development in the field.


**Figure 2: Word Cloud of High-Frequency Keywords in Linked Data Research (2000–2024)**

Figure 2 presents a word cloud visualization of the most frequently occurring keywords in linked data research over the past 25 years. The prominence of each term corresponds to its frequency of occurrence, highlighting the thematic concentration and evolving interests within the scholarly community.

The most dominant term is "semantic web" (3,019 occurrences), reflecting its foundational role in the field. Closely following are "linked data" (2,304) and "data handling" (2,068), which emphasize the technical and infrastructural backbone of linked data applications. Keywords such as "linked datum" (1,617) and "semantics" (1,356) underscore the conceptual and modeling aspects of knowledge representation.

Other prominent terms include "RDF" (994) and "ontology" (735), which indicate the centrality of semantic technologies and data structuring methods. Additionally, terms like "query processing", "SPARQL", and "metadata" highlight the operational layers of data access

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

and retrieval. The presence of keywords such as "open data", "world wide web", and "interoperability" point to broader trends of openness, integration, and usability.

This visualization not only reveals the core vocabulary of the field but also reflects the interdisciplinary nature of linked data research, spanning topics in computer science, artificial intelligence, and information management. It provides a concise thematic overview of the most influential and recurring concepts, thus guiding future research direction and contextual understanding.


**Figure 3: Treemap of High-Frequency Keywords in Linked Data Research (2000–2024)**

Figure 3 provides a Treemap visualization that depicts the relative frequency and importance of key terms used in linked data research publications over the past 25 years. Each rectangle's size corresponds to the proportional frequency of a term, offering a clear hierarchical view of dominant themes within the field.

- "Semantic web" occupies the largest area, with 3,019 occurrences (12%), reaffirming its foundational role in the domain.
- "Linked data" follows at 2,304 occurrences (9%), highlighting its prominence as both a concept and technological framework.
- "Data handling" (2,068; 8%) and "linked datum" (1,617; 6%) emphasize the technical and operational aspects.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

- Conceptual keywords like "semantics" (1,356; 5%) and "ontology" (735; 3%) illustrate the focus on meaning, structure, and classification.
- Technical standards and languages such as "RDF" (994; 4%), "SPARQL" (436; 2%), and "query processing" (576; 2%) are well represented, indicating active research on data querying and interoperability.

Secondary terms like "open data", "metadata", "linked open data", and "world wide web" each range from 2–3%, showing their influence in specific sub-domains. Other topics like "machine learning", "digital storage", "natural language processing", and "graph theory" also appear, reflecting interdisciplinary convergence.

This treemap succinctly visualizes the breadth and depth of thematic exploration in linked data research. By combining frequency and proportion, it allows for quick identification of core vs. emerging concepts, making it a valuable analytical tool for trend identification and future research direction.

**Table 13: Thematic Clustering by Coupling: Impact and Centrality of Research Themes**

| Label | Group | Freq | Centrality | Impact |
|---|---|---|---|---|
| Linked Data – Conf 66.7%, Semantic Web – Conf 70.5%, RDF – Conf 52.8% | 1 | 140 | 0.336 | 5.894 |
| Linked Data – Conf 33.3%, RDF – Conf 47.2%, Semantic Web – Conf 29.5% | 2 | 107 | 0.478 | 4.045 |

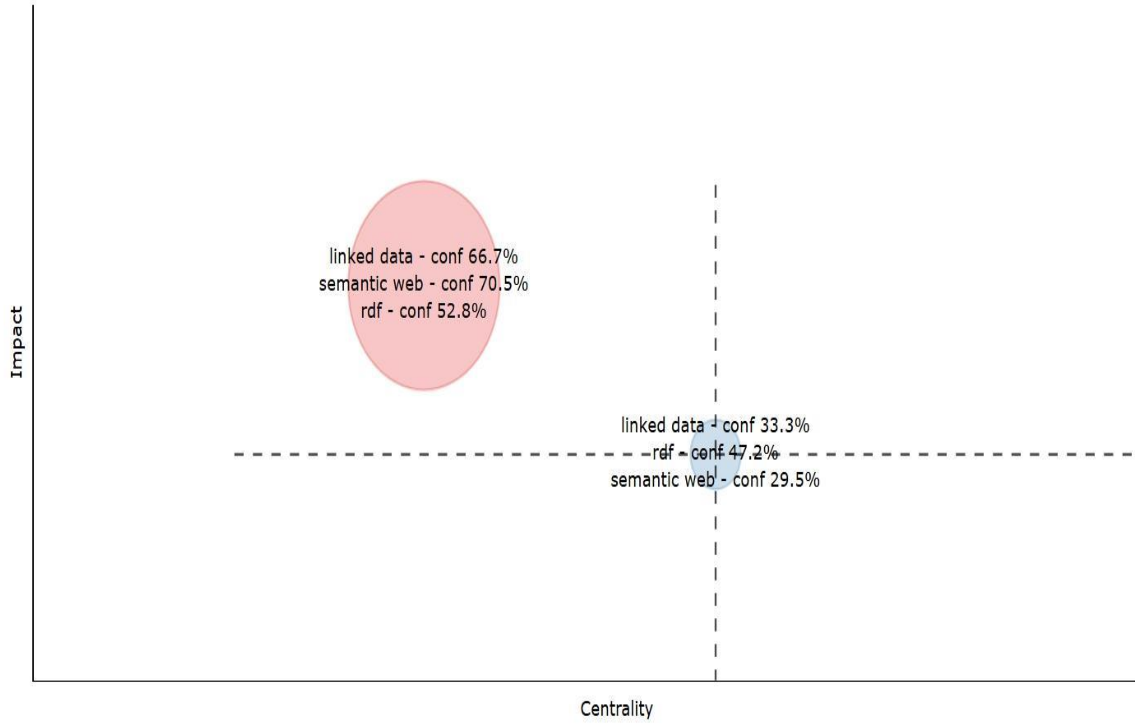


**Figure 4: Thematic Map of Coupling-Based Clustering in Linked Data Research (2000–2024)**

Table 13 and Figure 4 present the thematic clustering in linked data research based on bibliographic coupling, highlighting two major research themes over the period 2000–2024. These themes are analyzed in terms of their frequency, centrality, and impact. The first cluster, comprising “Linked Data – Conf 66.7%,” “Semantic Web – Conf 70.5%,” and “RDF – Conf

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

52.8%," demonstrates a relatively high impact score of 5.894 but a moderate centrality of 0.336. This indicates that while this cluster forms the intellectual and conceptual foundation of the field, its interconnection with other themes is somewhat limited. It reflects core theoretical and infrastructural developments within linked data scholarship.

In contrast, the second cluster, which includes "Linked Data – Conf 33.3%," "RDF – Conf 47.2%," and "Semantic Web – Conf 29.5%," exhibits a higher centrality of 0.478 and an impact score of 4.045. This suggests a more interdisciplinary role, indicating that this theme is more connected with other areas of research but slightly lower in citation impact. The placement of these clusters in the thematic map (Figure 4) reveals a strategic divergence in the field—where one cluster signifies influential foundational research and the other signifies emerging, well-connected applications. Thematic coupling analysis underscores the balanced development of the field across both depth (impact) and breadth (centrality).

**Table 14: Strategic Thematic Clusters in Linked Data Research**

| Cluster | CallonCentrality | CallonDensity | RankCentrality | RankDensity | ClusterFrequency |
|---|---|---|---|---|---|
| RDF | 0.703 | 3.068 | 2 | 4 | 9432 |
| Linked Data | 0.861 | 2.772 | 3 | 2 | 16363 |
| Semantic Web | 1.114 | 2.382 | 4 | 1 | 10908 |
| Human | 0.039 | 2.931 | 1 | 3 | 158 |

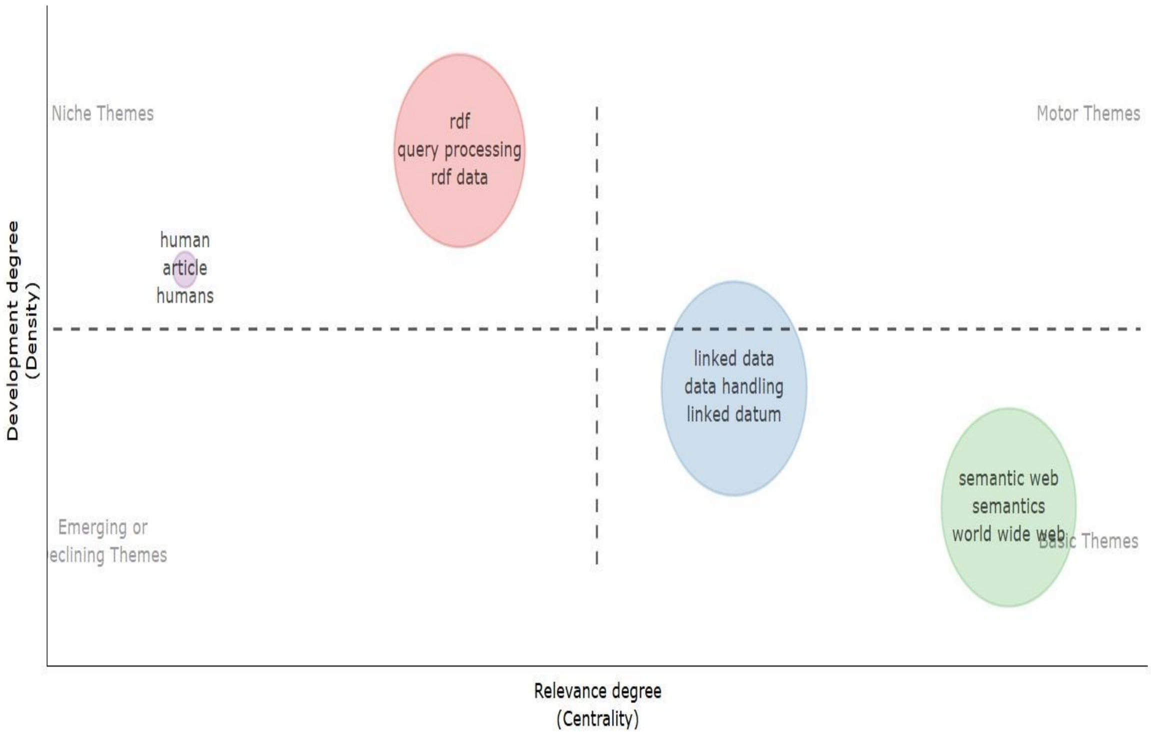


**Figure 5: Thematic Map Depicting Conceptual Structure Based on Centrality and Density**

Table 14 and Figure 5 illustrate the thematic structure of linked data research using a strategic thematic map. This map categorizes themes across two dimensions: Callon centrality (representing relevance or connectivity to other themes) and Callon density (indicating the level of internal development or cohesion within a theme).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

The semantic web cluster, with the highest centrality (1.114) and strong density (2.382), is positioned in the basic themes quadrant. This indicates that it plays a foundational role in the field, connecting with a wide range of other themes while maintaining moderate internal development. Similarly, the linked data cluster, which has the highest frequency (16,363) and notable centrality (0.861), appears in the motor themes quadrant. This suggests that it is a well-developed and highly integrated area, central to the field's intellectual structure and evolution.

In contrast, the RDF cluster appears in the niche themes quadrant with the highest density (3.068) and strong internal coherence, but slightly lower centrality (0.703), reflecting its focused and specialized nature. Meanwhile, the human cluster, which has the lowest centrality (0.039) but relatively high density (2.931), is placed in the emerging or declining themes quadrant, indicating limited connectivity with mainstream research yet some degree of internal cohesion.

This thematic mapping provides a strategic overview of the field's conceptual landscape, highlighting which clusters are central and enduring, and which are either emerging or potentially declining in influence.

**Table 15: Factorial Analysis of Highly Cited Articles by Cluster**

| Documents | dim1 | dim2 | contrib | TC | Cluster |
|---|---|---|---|---|---|
| Bizer, C. (2009). International Journal on Semantic Web and Information Systems | 0.02 | 0.02 | 0 | 3411 | 1 |
| Heath, T. (2011). Synthesis Lectures on the Semantic Web: Theory and Technology | -0.16 | -0.16 | 0 | 1199 | 1 |
| Horrocks, I. (2003). Web Semantics | -0.2 | -0.2 | 0 | 1161 | 1 |
| Broekstra, J. (2002). Lecture Notes in Computer Science | -0.46 | -0.46 | 0 | 841 | 1 |
| Nejdl, W. (2002). Proceedings of the International Conference on World Wide Web (WWW) | -0.23 | -0.23 | 0 | 516 | 1 |
| Zaveri, A. (2016). Semantic Web | 0.07 | 0.07 | 0 | 480 | 1 |
| Decker, S. (2000). IEEE Internet Computing | -0.18 | -0.18 | 0 | 464 | 1 |
| Neumann, T. (2010). VLDB Journal | -0.54 | -0.54 | 0 | 462 | 1 |
| Neumann, T. (2008). Proceedings of the VLDB Endowment | -0.32 | -0.32 | 0 | 461 | 1 |
| Unger, C. (2012). WWW – Proceedings of the Annual Conference on World Wide Web | -0.19 | -0.19 | 0 | 388 | 1 |

Table 15 presents the factorial analysis of the most highly cited documents in the linked data research domain, grouped under a single dominant cluster. All the listed articles fall within Cluster 1, indicating a cohesive intellectual structure among these foundational works. The coordinates on the two principal dimensions (Dim1 and Dim2) reflect their positioning in the conceptual space, although the low contribution scores suggest minimal individual influence on the overall variance in the factorial model. Notably, Bizer (2009) with 3411 citations and Heath (2011) with 1199 citations represent seminal contributions, anchoring the cluster thematically around linked data and the semantic web. Other influential works include Horrocks (2003), Broekstra (2002), and Neumann (2008, 2010), all of which have played key roles in advancing RDF data models, query optimization, and data integration. The clustering and citation impact emphasize the centrality of these documents in shaping the intellectual foundations and thematic evolution of the field over the past 25 years.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

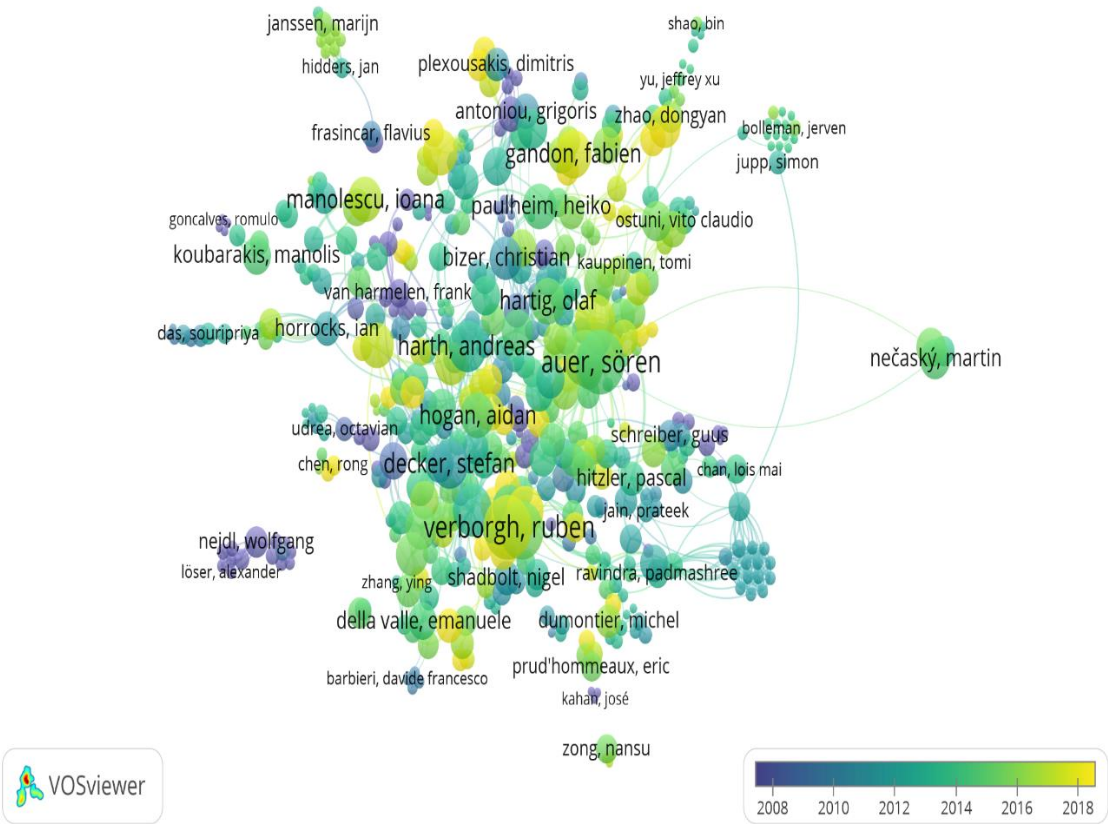


**Figure 6: Co-authorship Network Visualization Based on Authors with ≥100 Citations**

Figure 6 illustrates the co-authorship network among authors who have each received at least 100 citations in the field of linked data and semantic web research. The map, generated using VOSviewer, represents each author as a node, where the size of the node reflects their publication or citation count, and the thickness of the connecting lines denotes the strength of co-authorship ties.

The color gradient ranging from purple to yellow indicates the average publication year, with purple representing earlier years (circa 2008) and yellow indicating more recent activity (up to 2018).

Several prominent researchers, including Sören Auer, Christian Bizer, Heiko Paulheim, Ruben Verborgh, Olaf Hartig, and Andreas Harth, appear as large, centrally located nodes, signifying their influential role and extensive collaboration within the community. These individuals often serve as bridges between multiple research clusters, fostering interdisciplinary collaboration.

The network also reveals distinct clusters of authors, suggesting thematic or institutional alliances. In contrast, a few researchers, such as Martin Nečaský, appear isolated, indicating niche or less connected scholarly contributions. Visualization highlights a dense and interconnected research landscape with a clear evolution of author contributions over time.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

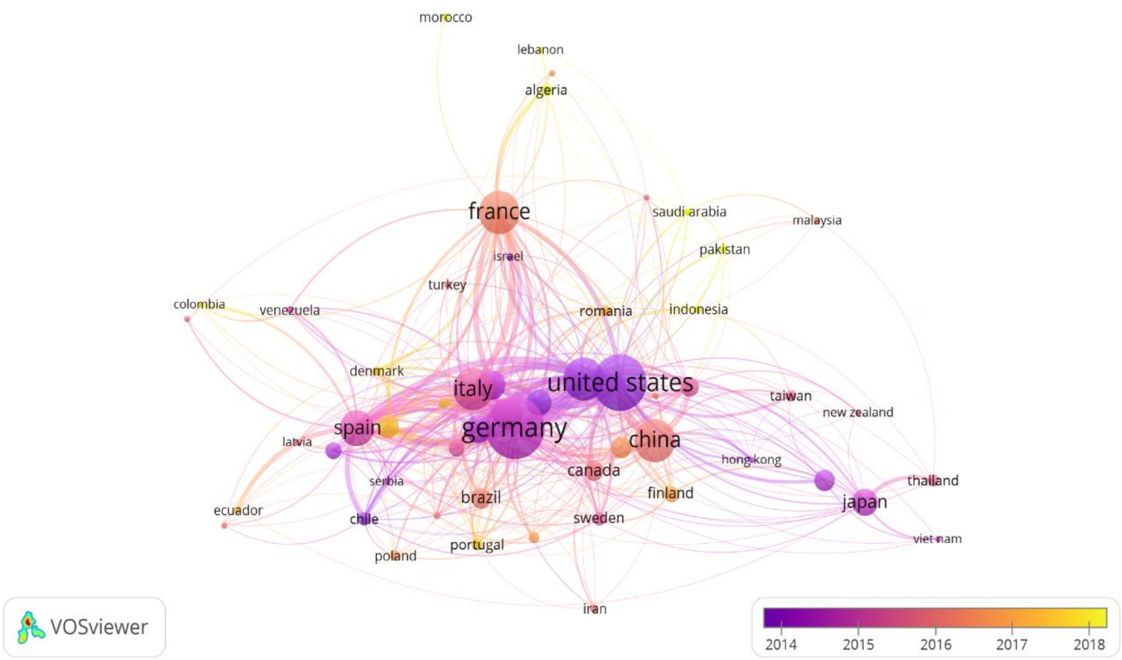


**Figure 7: Country-Level Co-authorship Network Visualization**

Figure 7 presents a co-authorship network map visualizing international collaborations among countries involved in linked data and semantic web research. The analysis was conducted in VOSviewer using co-authorship as the type of analysis, countries as the unit of analysis, and full counting as the counting method. Countries with at least one publication and a minimum of 100 citations were included, resulting in 56 qualifying countries. These countries were selected based on their total link strength—the sum of co-authorship links with other nations.

In the network, each node represents a country, and the size of the node indicates the number of documents produced. The thickness of the lines between nodes corresponds to the strength of collaborative ties, while the color gradient—from purple to yellow—depicts the average year of publication, with yellow indicating more recent contributions. Countries such as the United States, Germany, France, Italy, and China emerge as central players with high publication volume and dense co-authorship links, indicating their leading roles in global research collaboration.

The United States and Germany exhibit the highest total link strengths (463 and 615, respectively), reflecting their extensive partnerships. France, Italy, and the United Kingdom also demonstrate significant co-authorship activity, acting as hubs within their respective clusters. Countries such as India, Indonesia, and Vietnam, while present in the network, display relatively fewer collaborative links, suggesting either emerging involvement or less integration into dominant research networks. Interestingly, Venezuela, despite a modest publication count (27), shows an unusually high citation count (21,022), pointing to the high impact of specific works. The map illustrates a well-connected global research community with prominent collaboration among Western, Asian, and select Latin American countries. The distribution also highlights the gradual inclusion of researchers from diverse geographical backgrounds into the linked data research domain.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

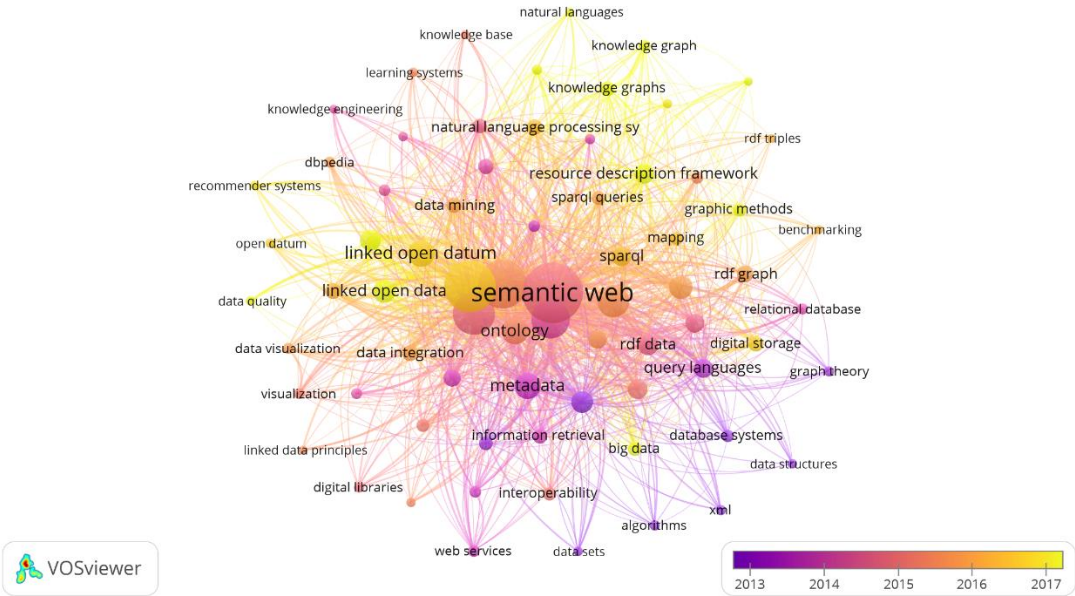


**Figure 8: Keyword Co-occurrence Network Visualization**

Figure 8 illustrates the co-occurrence network of high-frequency keywords (≥100 occurrences) derived from the dataset, revealing the thematic structure of research in the domain of linked data and semantic web technologies. The visualization was created using VOSviewer, where each node represents a keyword, and its size corresponds to the number of times that keyword appeared across documents. The color of each node indicates the average year of publication associated with that keyword, following a spectrum from purple (earlier) to yellow (more recent). The strength and thickness of links between nodes reflect the frequency with which pairs of keywords co-occurred, providing insight into conceptual relationships.

The term "semantic web" is the most dominant node in the network, showing the highest number of occurrences (3017) and total link strength (12,847), firmly establishing it as the core research theme. Closely surrounding it are keywords such as "linked data" (2304 occurrences), "linked open data" (534), "metadata" (708), "ontology" (735), and "SPARQL" (435), which highlight key pillars of the semantic web ecosystem. Keywords like "natural language processing systems", "artificial intelligence", and "knowledge graphs" indicate intersections with emerging technologies and suggest the increasing influence of machine learning and AI on semantic technologies.

Clusters within the network reveal coherent topical groups—for example, one cluster centers around data representation and querying (e.g., RDF, SPARQL, RDF triples), while another includes interoperability, metadata, and data integration, indicating applications in digital libraries and knowledge management. The visualization also captures growing interests in newer topics such as "knowledge graph", "big data", and "user interfaces", reflected in lighter yellow shades corresponding to recent publication years (2016–2017).

This keyword co-occurrence map identifies the foundational and evolving themes within the field. It demonstrates the centrality of semantic web technologies while highlighting how adjacent fields such as data mining, artificial intelligence, and data visualization are shaping ongoing research directions.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

## 5. DISCUSSION

The bibliometric and thematic analysis of Linked Data (LD) research from 2000 to 2024 reveals a field that has matured both conceptually and methodologically, with strong disciplinary integration in library and information science, computer science, and the broader open science ecosystem. The findings align with the core objectives of this study, offering comprehensive insights into scholarly trends, intellectual structures, and practical applications of LD technologies.

### 5.1 Evolution of Scholarly Output and Intellectual Influence

The growth of LD publications (Table 2) indicates a robust and sustained scholarly interest, particularly between 2010 and 2016, when publication volume peaked. This aligns with key technological and institutional developments, such as the launch of BIBFRAME (Fortier et al., 2022) and the proliferation of Linked Data initiatives like Europeana and LD4L. The decline in publications after 2016 may signal a saturation in foundational topics or a pivot toward domain-specific applications and integrations with other technologies (e.g., AI, NLP), as reflected in thematic mappings and cluster analyses (Hosseini et al., 2025).

Citation data (Table 3) reinforces the continued influence of early foundational works, particularly Bizer et al. (2009), whose conceptualization of LD remains central to ongoing scholarship. These early works have not only set the theoretical tone for the domain but also guided practical implementations in libraries and open data infrastructures (Wu & Ye, 2021).

### 5.2 Conceptual and Thematic Evolution

Thematic and co-word analyses (Figure 4, Table 14) reveal a field organized around foundational themes such as RDF, SPARQL, and the Semantic Web, with emerging overlaps into deep learning, natural language processing, and knowledge representation (Hosseini et al., 2025). This convergence underscores a gradual shift from technical standardization toward intelligent applications of LD in computational and user-centric domains.

The strong presence of themes like metadata interoperability and FAIR data implementation confirms LD's increasing relevance in structuring digital knowledge ecosystems (Wilkinson et al., 2016; Bhat & Wani, 2025). Applications in libraries (Gaitanou et al., 2024) reflect this evolution, especially through hybrid models (e.g., BIBFRAME and IFLA LRM) that facilitate semantic enrichment and cross-system linking (Alemu et al., 2012).

### 5.3 Institutional and Geographical Collaboration

The co-authorship and country network analyses (Figures 6 and 7; Tables 9 and 10) illustrate a well-connected, international research community, particularly concentrated in Europe, China, and North America. Germany and the Netherlands, in particular, stand out for their high average citation rates, suggesting that collaborative outputs are not only prolific but also impactful (Gupta et al., 2020).

Despite China's high publication volume, its relatively low citation-per-paper average may reflect challenges in visibility or alignment with dominant Western frameworks (Wahid et al., 2018). Conversely, countries like Ireland and the Netherlands demonstrate high-impact, lower-volume contributions, consistent with the FAIR and open data leadership in the European research policy landscape (Mons et al., 2017).

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

### 5.4 Library Applications and Metadata Transformation

The bibliometric results confirm that LD's most tangible impact lies in libraries and cultural heritage institutions. Projects like BIBFRAME, Europeana EDM, and Wikidata integration represent a paradigm shift from legacy metadata practices to open, machine-readable, semantically rich frameworks (Kroeger, 2013; Haslhofer et al., 2019).

However, the persistence of legacy systems such as MARC, along with challenges in vocabulary alignment and metadata quality, continue to obstruct full-scale implementation (Zaveri et al., 2015; Wahid et al., 2018). This resonates with Gaitanou et al. (2024), who note that many LD-based systems in libraries remain in pilot or transitional phases, signaling the need for sustained institutional investment, skill development, and standards harmonization.

### 5.5 Authority Control and Knowledge Graph Integration

The integration of Linked Data into authority control frameworks—such as VIAF, ISNI, and Wikidata—has significantly improved entity disambiguation and enriched cataloging practices (Zhu, 2019; Wiederhold & Reeve, 2021). Bibliometric cluster analysis (Table 13) underscores the centrality of these themes, reflecting their foundational role in bridging bibliographic control and open knowledge infrastructures.

The deployment of Knowledge Graphs in libraries, as shown in the work of Clark et al. (2022) and Lu & Jimei (2024), is accelerating this transformation. These systems support semantic linking, personalized content delivery, and intelligent search interfaces, extending the utility of LD beyond cataloging to dynamic user interaction and data discovery (Lüschow, 2022; Lytras et al., 2019).

### 5.6 Linked Data in Open Science and FAIR Ecosystems

The convergence of LD with FAIR principles demonstrates its critical role in scientific data stewardship. High-frequency keywords (Figure 8) such as "interoperability," "metadata," and "FAIR data" point to growing alignment with open science objectives (Wilkinson et al., 2016; Frey & Hellmann, 2021). The European Open Science Cloud (EOSC) and related infrastructures provide exemplars of FAIR-compliant Linked Data implementation at scale (Larsson et al., 2025).

Despite this progress, practical challenges persist, including semantic inconsistencies, data conversion barriers, and limited adoption among smaller institutions (Singh & Maurya, 2025). Continued efforts in developing metadata validation tools (e.g., YAMA) and semantic crosswalk frameworks (Thalhath et al., 2025) are essential for achieving scalable and inclusive FAIR ecosystems.

## 6. CONCLUSION

This bibliometric and thematic analysis of Linked Data (LD) research from 2000 to 2024 reveals a field that has evolved from conceptual foundations into a multifaceted domain of practical, disciplinary, and cross-disciplinary applications. The sustained growth in publication volume and international collaboration underscores the increasing academic and institutional relevance of LD technologies, particularly in libraries, digital knowledge systems, and open science infrastructures.

The results confirm that libraries and cultural heritage institutions are at the forefront of LD adoption, transitioning from isolated metadata silos to semantically rich, interoperable

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

frameworks through initiatives like BIBFRAME, EDM, and Wikidata integration. These developments, however, remain uneven, with many implementations still in pilot phases due to technical complexity, legacy systems, and a lack of standardization and institutional readiness.

Thematically, the field has expanded beyond foundational technologies such as RDF and SPARQL to encompass emerging areas like knowledge graphs, natural language processing, and AI-enhanced metadata curation. This convergence signals a shift toward intelligent, user-centered applications of Linked Data in scholarly communication and library systems. The alignment with FAIR data principles further emphasizes LD's growing role in facilitating transparency, interoperability, and reuse of scientific data across disciplines.

Nevertheless, several challenges persist, including inconsistencies in metadata vocabularies, barriers to converting legacy records, and gaps in semantic alignment. These findings suggest that while LD offers powerful solutions for bibliographic control and data integration, realizing its full potential will require sustained efforts in tool development, training, and institutional policy alignment.

This study provides a comprehensive empirical foundation for understanding the trajectory of Linked Data research. It invites future inquiry into domain-specific implementations, user-centered design, and the role of Linked Data in emerging paradigms such as data justice, ethical AI, and multilingual knowledge representation.

### 6.1 Implications for Practice

The findings of this study highlight the critical role of libraries, archives, and cultural heritage institutions as early adopters and promoters of Linked Data technologies. For practitioners, the results underscore the need to embed Linked Data principles into day-to-day cataloging, metadata enrichment, and digital repository management workflows. Moving beyond pilot projects, library professionals should be equipped to manage RDF-based records, integrate external knowledge bases such as Wikidata and VIAF, and maintain persistent identifiers to ensure resource findability. The adoption of hybrid models, such as BIBFRAME in conjunction with IFLA LRM, offers a practical pathway for transitioning from MARC-based systems while preserving interoperability with legacy infrastructure. Furthermore, Linked Data can enhance public service delivery by enabling user-centric discovery layers, semantic search capabilities, and intelligent recommendation systems.

### 6.2 Policy Recommendations

To support sustained Linked Data adoption, institutional and policy frameworks must prioritize capacity building, vocabulary harmonization, and FAIR-aligned metadata governance. First, national and regional library networks should invest in ongoing professional development programs that train catalogers, metadata specialists, and IT staff in Semantic Web standards, ontology management, and Linked Data publishing. Second, cross-institutional agreements on controlled vocabularies, schema mappings, and metadata application profiles are essential to avoid semantic inconsistencies and ensure cross-platform interoperability. Third, policymakers in cultural heritage and research sectors should mandate FAIR-aligned metadata practices—such as the use of persistent identifiers, open licensing, and machine-actionable formats—as part of funding and repository accreditation requirements. Finally, collaboration with international standard bodies (e.g., W3C, IFLA) will help align local implementations with global best practices, fostering a truly interconnected and reusable web of cultural and scholarly data.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

**References**


1) Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. "Linked Data for Libraries." *New Library World* 113, no. 11/12 (2012): 549–70. https://doi.org/10.1108/03074801211282920.
2) Bhat, Ajra, and Zahid Ashraf Wani. "The FAIRification Process for Data Stewardship: A Comprehensive Discourse on the Implementation of the FAIR Principles for Data Visibility, Interoperability and Management." *IFLA Journal* 51, no. 2 (2025): 382–98. https://doi.org/10.1177/03400352241270692.
3) Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far:" *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 1–22. https://doi.org/10.4018/jswis.2009081901.
4) Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far." In *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*, 1st ed., vol. 52. Association for Computing Machinery, 2023. https://doi.org/10.1145/3591366.3591378.
5) Candela, Gustavo, Pilar Escobar, Rafael C Carrasco, and Manuel Marco-Such. "Evaluating the Quality of Linked Open Data in Digital Libraries." *Journal of Information Science* 48, no. 1 (2022): 21–43. https://doi.org/10.1177/0165551520930951.
6) Capurro, Carlotta, and Gertjan Plets. "Europeana, EDM, and the Europeanisation of Cultural Heritage Institutions." *Digital Culture & Society* 6, no. 2 (2020): 163–90. https://doi.org/10.14361/dcs-2020-0209.
7) Clark, Jason A., Helen K.R. Williams, and Doralyn Rossmann. "Wikidata and Knowledge Graphs in Practice: Using Semantic SEO to Create Discoverable, Accessible, Machine-Readable Definitions of the People, Places, and Services in Libraries and Archives." *Information Services and Use* 42, nos. 3–4 (2022): 377–90. https://doi.org/10.3233/ISU-220171.
8) Cole, Timothy W., Myung-Ja Han, William Fletcher Weathers, and Eric Joyner. "Library Marc Records into Linked Open Data: Challenges and Opportunities." *Journal of Library Metadata* 13, nos. 2–3 (2013): 163–96. https://doi.org/10.1080/19386389.2013.826074.
9) Costa, Tiago, Teresa Borges-Tiago, Francisco Martins, and Flávio Tiago. "System Interoperability and Data Linkage in the Era of Health Information Management: A Bibliometric Analysis." *Health Information Management Journal*, September 16, 2024, 18333583241277952. https://doi.org/10.1177/18333583241277952.
10) Danskin, Alan. "Linked and open data : RDA and bibliographic control." *JLIS : Italian Journal of Library, Archives and Information Science = Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione* 4, no. 1 (2013): 147–60. https://doi.org/10.4403/jlis.it-5463.
11) Doerr, Martin, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and Herbert van de Sompel. "The Europeana Data Model (EDM)." IFLA World Library and Information Congress, 2010, 12. https://www.researchgate.net/publication/303058300_The_Europeana_Data_Model_EDM.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918


12) Dunning, Alastair, Madeleine de Smaele, and Jasmin Böhmer. "Are the FAIR Data Principles Fair?" *International Journal of Digital Curation* 12, no. 2 (2017): 2. https://doi.org/10.2218/ijdc.v12i2.567.

13) Dutta, Biswanath, Nabina Khamaru, and Harikrishnan Js. "Linked Data Adaptation and Practice in Libraries: An Indian Panorama." *The Electronic Library* 43, no. 2 (2025): 132–53. https://doi.org/10.1108/EL-06-2024-0167.

14) Fortier, Alexandre, Heather J. Pretty, and Daniel B. Scott. "Assessing the Readiness for and Knowledge of BIBFRAME in Canadian Libraries." *Cataloging & Classification Quarterly* 60, no. 8 (2022): 708–35. https://doi.org/10.1080/01639374.2022.2119456.

15) Frey, Johannes, and Sebastian Hellmann. "FAIR Linked Data - Towards a Linked Data Backbone for Users and Machines." *Companion Proceedings of the Web Conference 2021* (New York, NY, USA), WWW '21, Association for Computing Machinery, June 3, 2021, 431–35. https://doi.org/10.1145/3442442.3451364.

16) Gaitanou, Panorea, Ioanna Andreou, Miguel-Angel Sicilia, and Emmanouel Garoufallou. "Linked Data for Libraries: Creating a Global Knowledge Space, a Systematic Literature Review." *Journal of Information Science* 50, no. 1 (2024): 204–44. https://doi.org/10.1177/01655515221084645.

17) Gingras, Yves. *Bibliometrics and Research Evaluation: Uses and Abuses*. The MIT Press, 2016. https://doi.org/10.7551/mitpress/10719.001.0001.

18) Gunaratna, Kalpa, Sarasi Lalithsena, and Amit Sheth. "Alignment and Dataset Identification of Linked Data in Semantic Web." *WIREs Data Mining and Knowledge Discovery* 4, no. 2 (2014): 139–51. https://doi.org/10.1002/widm.1121.

19) Gupta, Brij Mohan, S.M. Dhawan, Neeraj Singh, and Ashok Kumar. "Linked Data: A Scientometrics Assessment of Global Publications Output during 1996–2019." *International Journal of Information Dissemination and Technology* 10, no. 1 (2020): 62–69. https://doi.org/10.5958/2249-5576.2020.00010.2.

20) Hakimov, Sherzod, Salih Atilay Oto, and Erdogan Dogdu. "Named Entity Recognition and Disambiguation Using Linked Data and Graph-Based Centrality Scoring." *Proceedings of the 4th International Workshop on Semantic Web Information Management* (New York, NY, USA), SWIM '12, Association for Computing Machinery, May 20, 2012, 1–7. https://doi.org/10.1145/2237867.2237871.

21) Haller, Armin, Javier D. Fernández, Maulik R. Kamdar, and Axel Polleres. "What Are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web." *J. Data and Information Quality* 12, no. 2 (2020): 9:1-9:34. https://doi.org/10.1145/3369875.

22) Hallo, María, Sergio Luján-Mora, Alejandro Maté, and Juan Trujillo. "Current State of Linked Data in Digital Libraries." *Journal of Information Science* 42, no. 2 (2016): 117–27. https://doi.org/10.1177/0165551515594729.

23) Haslhofer, Bernhard, Antoine Isaac, and Rainer Simon. "Knowledge Graphs in the Libraries and Digital Humanities Domain." In *Encyclopedia of Big Data Technologies*, edited by Sherif Sakr and Albert Y. Zomaya. Springer International Publishing, 2019. https://doi.org/10.1007/978-3-319-77525-8_291.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918


24) Hasnain, Ali, and Dietrich Rebholz-Schuhmann. “Assessing FAIR Data Principles Against the 5-Star Open Data Principles.” In *The Semantic Web: ESWC 2018 Satellite Events*, edited by Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, et al. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-98192-5_60.

25) Heng, Greta, Patricia Lampron, and Myung-Ja Han. “Linked Data Editors: Where Are We Now, Challenges and Considerations.” *Library Resources & Technical Services* 68, no. 3 (2024): 3. https://doi.org/10.5860/lrts.68n3.8226.

26) Hitzler, Pascal, and Krzysztof Janowicz. “Linked Data, Big Data, and the 4th Paradigm.” *Semantic Web* 4, no. 3 (2013): 233–35. https://doi.org/10.3233/SW-130117.

27) Hogan, Aidan. “Linked Data & the Semantic Web Standards.” In *Linked Data Management*. Chapman and Hall/CRC, 2014.

28) Hosseini, Elaheh, Amir Ghaebi, and Roya Baradar. “Bibliometrics and Mapping of Co-Words in the Field of Linked Data.” *Scientometrics Research Journal* 7, no. (Issue 1, spring&summer) (2021): 91–116. https://doi.org/10.22070/rsci.2020.4904.1333.

29) Hosseini, Elaheh, Kimiya Taghizadeh Milani, and Mohammad Shaker Sabetnasab. “Development and Maturity of Co-Word Thematic Clusters: The Field of Linked Data.” *Library Hi Tech* 43, no. 1 (2025): 81–113. https://doi.org/10.1108/lht-10-2022-0488.

30) Huang, Hong, and Jian Qin. “Metadata Functional Requirements for Genomic Data Practice and Curation.” *Information Research an International Electronic Journal* 29, no. 2 (2024): 2. https://doi.org/10.47989/ir292363.

31) Jacob, Daniel, François Ehrenmann, Romain David, Joseph Tran, Cathleen Mirande-Ney, and Philippe Chaumeil. “An Ecosystem for Producing and Sharing Metadata within the Web of FAIR Data.” *GigaScience* 14 (January 2025): giae111. https://doi.org/10.1093/gigascience/giae111.

32) Jia, Junzhi. “From Data to Knowledge: The Relationships between Vocabularies, Linked Data and Knowledge Graphs.” *Journal of Documentation* 77, no. 1 (2020): 93–105. World. https://doi.org/10.1108/JD-03-2020-0036.

33) Jin, Qiang, Jim Hahn, and Gretchen Croll. “BIBFRAME Transformation for Enhanced Discovery.” *Library Resources & Technical Services* 60, no. 4 (2016): 4. https://doi.org/10.5860/lrts.60n4.223.

34) Kamal, Ahmad M., and Koraljka Golub. *Subject Matters : Metadata Standards and Subject Access for Library and Museum Catalogues*. National Library of Norway, 2025. https://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-136374.

35) Kroeger, Angela. “The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future.” *Cataloging & Classification Quarterly* 51, no. 8 (2013): 873–90. https://doi.org/10.1080/01639374.2013.823584.

36) Larsson, Åsa M., Barbro Bornsäter, and Marei Hacke. “Developing Practices for FAIR and Linked Data in Heritage Science.” *Npj Heritage Science* 13, no. 1 (2025): 1–13. https://doi.org/10.1038/s40494-025-01598-x.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918


37) Lauscher, Anne, Kai Eckert, Lukas Galke, et al. "Linked Open Citation Database: Enabling Libraries to Contribute to an Open and Interconnected Citation Graph." *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, ACM, May 23, 2018, 109–18. https://doi.org/10.1145/3197026.3197050.

38) Leiva-Mederos, Amed, José A. Senso, Sandor Domínguez-Velasco, and Pedro Hípola. "AUTHORIS: A Tool for Authority Control in the Semantic Web." *Library Hi Tech* 31, no. 3 (2013): 536–53. world. https://doi.org/10.1108/LHT-12-20112-0135.

39) Li, Kai, Brian Dobreski, and Molly Busch. "How Are Library Cataloging Metadata Used Differently over Time: A Large-Scale Quantitative Analysis of the Library of Congress Catalog." *Journal of Documentation* 81, no. 2 (2025): 385–402. world. https://doi.org/10.1108/JD-08-2024-0199.

40) Lu, Chen, and Zhu Jimei. "Visualizing the Construction of Subject Librarianship in Chinese Higher Education Libraries: A Knowledge Graph-Based Analysis." *Journal of Librarianship and Information Science*, June 25, 2024, 09610006241259494. https://doi.org/10.1177/09610006241259494.

41) Lüschow, Andreas. "Application of Graph Theory in the Library Domain—Building a Faceted Framework Based on a Literature Review." *Journal of Librarianship and Information Science* 54, no. 4 (2022): 558–77. https://doi.org/10.1177/09610006211036734.

42) Lytras, Miltiadis D., Saeed-Ul Hassan, and Naif Radi Aljohani. "Linked Open Data of Bibliometric Networks: Analytics Research for Personalized Library Services." *Library Hi Tech* 37, no. 1 (2019): 2–7. https://doi.org/10.1108/lht-03-2019-277.

43) McCallum, Sally. "BIBFRAME Development." *JLIS : Italian Journal of Library, Archives and Information Science* 8, no. 3 (2017): 71–85. https://doi.org/10.4403/jlis.it-12415.

44) McKenna, Lucy, Christophe Debruyne, and Declan O'Sullivan. "Using Linked Data to Create Provenance-Rich Metadata Interlinks: The Design and Evaluation of the NAISC-L Interlinking Framework for Libraries, Archives and Museums." *AI & SOCIETY* 37, no. 3 (2022): 921–47. https://doi.org/10.1007/s00146-021-01373-z.

45) Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, and Mark D. Wilkinson. "Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud." *Information Services and Use* 37, no. 1 (2017): 49–56. https://doi.org/10.3233/ISU-170824.

46) Mountantonakis, Michalis, and Yannis Tzitzikas. "Large-Scale Semantic Integration of Linked Data: A Survey." *ACM Comput. Surv.* 52, no. 5 (2019): 103:1-103:40. https://doi.org/10.1145/3345551.

47) Myntti, Jeremy, and Nate Cothran. "Authority Control in a Digital Repository: Preparing for Linked Data." *Journal of Library Metadata* 13, nos. 2–3 (2013): 95–113. https://doi.org/10.1080/19386389.2013.826061.

48) Narock, Tom, and Hayden Wimmer. "Linked Data Scientometrics in Semantic E-Science." *Computers & Geosciences* 100 (March 2017): 87–93. https://doi.org/10.1016/j.cageo.2016.12.008.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918


49) Peroni, Silvio, Francesca Tomasi, and Fabio Vitali. "Reflecting on the Europeana Data Model." In *Digital Libraries and Archives*, edited by Maristella Agosti, Floriana Esposito, Stefano Ferilli, and Nicola Ferro. Springer, 2013. https://doi.org/10.1007/978-3-642-35834-0_23.

50) Quarati, Alfonso, and Riccardo Albertoni. "Linked Open Government Data: Still a Viable Option for Sharing and Integrating Public Data?" *Future Internet* 16, no. 3 (2024): 3. https://doi.org/10.3390/fi16030099.

51) Řezník, Tomáš, Lieven Raes, Andrew Stott, et al. "Improving the Documentation and Findability of Data Services and Repositories: A Review of (Meta) Data Management Approaches." *Computers & Geosciences* 169 (December 2022): 105194. https://doi.org/10.1016/j.cageo.2022.105194.

52) Saleem, Qurat Ul Ain, Nadia Butt, and Nosheen Fatima Warriach. "Applications of Linked Data Technologies in Libraries: Technical and Ethical Considerations." *2018 International Conference on Information Management and Processing (ICIMP)*, IEEE, January 2018, 11–15. https://doi.org/10.1109/ICIMP1.2018.8325833.

53) Shah, Naimat Ullah, Salman Bin Naeem, and Robina Bhatti. "Digital Data Sets Management in University Libraries: Challenges and Opportunities." *Global Knowledge, Memory and Communication* 74, no. 1/2 (2025): 446–62. world. https://doi.org/10.1108/GKMC-06-2022-0150.

54) Singh, Gaurav, and Adarsh Maurya. "The Role of Metadata in Data Curation for Enhancing Discoverability in Large Datasets." *International Journal of Web of Multidisciplinary Studies* 2, no. 1 (2025): 1.

55) Soiland-Reyes, Stian, Carole Goble, and Paul Groth. "Evaluating FAIR Digital Object and Linked Data as Distributed Object Systems." *PeerJ Computer Science* 10 (April 2024): e1781. https://doi.org/10.7717/peerj-cs.1781.

56) Thalhath, Nishad, Mitsuharu Nagamori, and Tetsuo Sakaguchi. "Metadata Application Profile as a Mechanism for Semantic Interoperability in FAIR and Open Data Publishing." *Data and Information Management* 9, no. 1 (2025): 100068. https://doi.org/10.1016/j.dim.2024.100068.

57) Ullah, Irfan, Shah Khusro, Asim Ullah, and Muhammad Naeem. "An Overview of the Current State of Linked and Open Data in Cataloging." *Information Technology and Libraries* 37, no. 4 (2018): 47–80. https://doi.org/10.6017/ital.v37i4.10432.

58) Umbach, Gaby. "Open Science and the Impact of Open Access, Open Data, and FAIR Publishing Principles on Data-Driven Academic Research: Towards Ever More Transparent, Accessible, and Reproducible Academic Output?" *Statistical Journal of the IAOS* 40, no. 1 (2024): 59–70. https://doi.org/10.3233/SJI-240021.

59) W3C. "Library Linked Data Incubator Group Final Report." 2011. https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/.

60) Wahid, Nazia, Nosheen Fatima Warraich, and Muzammil Tahira. "Mapping the Cataloguing Practices in Information Environment: A Review of Linked Data Challenges." *Information and Learning Science* 119, no. 9/10 (2018): 586–96. https://doi.org/10.1108/ils-10-2017-0106.

Volume 64 | Issue 12 | December 2025
DOI: 10.5281/zenodo.18082918

61) Warraich, Nosheen Fatima, and Abebe Rorissa. "Impediments to and Readiness for Linked Data Application in Libraries: Pakistani Information Professionals' Perspective." *Journal of Librarianship and Information Science* 56, no. 2 (2024): 341–52. https://doi.org/10.1177/09610006221141690.

62) Wiederhold, Rebecca A., and Gregory F. Reeve. "Authority Control Today: Principles, Practices, and Trends." In *Cataloging and Classification*. Routledge, 2021.

63) Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, no. 1 (2016): 160018. https://doi.org/10.1038/sdata.2016.18.

64) Wilkinson, Sean R., Meznah Aloqalaa, Khalid Belhajjame, et al. "Applying the FAIR Principles to Computational Workflows." *Scientific Data* 12, no. 1 (2025): 328. https://doi.org/10.1038/s41597-025-04451-9.

65) Wu, Anjie, and Yong Ye. "Bibliometric Analysis on Bibliometric-Based Ontology Research." *Science & Technology Libraries* 40, no. 4 (2021): 435–53. https://doi.org/10.1080/0194262x.2021.1920555.

66) Zhang, Lu. "Describe Library Resources with Knowledge Graph." IFLA World Library and Information Congress, 2017, 9. https://library.ifla.org/id/eprint/2753/.

67) Zhang, Yun, Weina Hua, and Shunbo Yuan. "Mapping the Scientific Research on Open Data: A Bibliometric Review." *Learned Publishing* 31, no. 2 (2018): 95–106. https://doi.org/10.1002/leap.1110.

68) Zhu, Lihong. "The Future of Authority Control: Issues and Trends in the Linked Data Environment." *Journal of Library Metadata* 19, nos. 3–4 (2019): 215–38. https://doi.org/10.1080/19386389.2019.1688368.

## APPENDIX-A

**Search Query:**

The dataset for this study was retrieved from the Scopus database using a carefully constructed search query to identify relevant research on Linked Data. The final search string TITLE ( "Linked Data" OR "Linked Open Data" OR "RDF" OR "RDF Data" OR "RDF Triples" OR "Linked Data Technologies" OR "Linked Data Cloud" OR "Linked Data Applications" OR "Linked Data Publishing" OR "Linked Data Integration" OR "Linked Data Framework" OR "Linked Data Querying" OR "Bibliographic Linked Data" OR "Library Linked Data" OR "Linked Library Data" OR "Catalog Linked Data" OR "Authority Linked Data" OR "Metadata Linked Data" OR "Library Ontology" OR "BIBFRAME" OR "SKOS" OR "Semantic Cataloging" OR "Semantic Metadata" OR "FAIR Linked Data" OR "LOD in Libraries" OR "LOD in Digital Libraries" OR "Knowledge Organization and Linked Data" OR "Cultural Heritage Linked Data" OR "Archival Linked Data" OR "Museum Linked Data" OR "Semantic Web in Libraries" OR "Library Knowledge Graphs" OR "Metadata Interoperability" OR "Digital Library Ontologies" OR "Library Semantic Interoperability" ) AND PUBYEAR > 1999 AND PUBYEAR < 2025 AND ( LIMIT-TO ( SUBJAREA , "COMP" ) OR LIMIT-TO ( SUBJAREA , "SOCI" ) ) AND ( LIMIT-TO ( DOCTYPE , "cp" ) OR LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( PUBSTAGE , "final" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ).