

Data Stewardship Research Trends in Organizations

Loso Judijanto
IPOSS Jakarta, Indonesia

Article Info

Article history:

Received Aug, 2026
Revised Aug, 2026
Accepted Aug, 2026

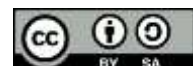
Keywords:

Data Stewardship
Data Governance
Data Management
Data Quality
FAIR Principles
Bibliometric Analysis

ABSTRACT

Data stewardship has become an increasingly important research area as organizations recognize data as a strategic asset requiring effective governance, quality management, and responsible utilization. This study aims to examine the development, intellectual structure, and emerging trends of data stewardship research in organizational contexts through a bibliometric analysis approach. Data were collected from the Scopus database using relevant keywords related to data stewardship and organizations, followed by performance analysis and science mapping using VOSviewer. The findings reveal that data stewardship research is strongly influenced by foundational studies on the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, data governance, and research data management. Keyword analysis indicates that data stewardship is closely associated with data management, metadata, data quality, information management, data sharing, privacy, and artificial intelligence. The temporal analysis shows a shift from traditional data management practices toward broader organizational capabilities involving ethical data governance, responsible data sharing, and digital transformation. Collaboration analysis demonstrates that research contributions are concentrated in leading countries and institutions, particularly in the United States and healthcare-related research environments, while international collaboration continues to expand. This study contributes by providing a comprehensive overview of the intellectual development of data stewardship research and identifying future opportunities for integrating governance, technology, and organizational capability perspectives.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Name: Loso Judijanto
Institution: IPOSS Jakarta, Indonesia
Email: losojudijantobumn@gmail.com

1. INTRODUCTION

The quick pace of development in digital technologies has turned data from an inert component of organizational activities into an asset that affects the processes of decision making, innovation, performance efficiency, and competitiveness. In order to enhance their performance and create

business value, companies from all types of industries actively use the methods of large-scale data production, analytics, artificial intelligence, and digital platforms. At the same time, the rise of data volume, variety, and complexity within organizations has led to the emergence of new problems related to data quality, availability, ownership, security, and proper use of information resources [1],

[2]. The lack of proper mechanisms of data management results in poor decision-making, regulatory issues, inefficiency, and mistrust within organizational environments [3], [4].

In such a setting, the concept of data governance has gained considerable relevance as a tool which ensures systematic approach and alignment with the strategic priorities of an organization regarding data management. Data governance can be understood as the set of decision rights, accountabilities, policies, standards, and processes that ensure proper acquisition, management, sharing, protection, and use of data assets in an organization [5], [6]. Despite the general framework offered by data governance, its effective functioning is closely linked to people who put this framework into practice in their day-to-day operations. In particular, one of the most prominent roles related to such operation is that of a data steward. Data stewardship can be defined as the process which implies personal accountability and authority for the management of the data quality and usage, compliance, and coordination between technical and business parties [6].

Data stewardship has gone past the typical notion involving data management and data documentation processes. Data stewards have become essential figures who enhance data quality, data literacy, metadata management, compliance with regulations, and ethical use of data. As more companies experience concerns such as privacy issues, artificial intelligence, and data-based decisions, data stewards play a vital role in developing organizational trust of data resources [7], [8]. Additionally, data stewardship entails a socio-technical approach since good stewardship involves interactions between technologies, organizational structures, and human capabilities. This approach indicates that data issues can be managed not only by using technology but also by creating proper organizational structures and mechanisms of cooperation between data producers and consumers.

The rising significance of data stewardship has attracted scholarly attention

within diverse fields, such as information systems, knowledge management, governance studies, and digital transformation studies. Scholars have explored data stewardship from diverse angles, including data quality management, information governance in enterprises, accountability mechanisms, and regulation compliance. For instance, scholars have analyzed how data stewardship functions can contribute to data governance maturity, information quality, and the organizational ability to manage digital assets [9], [10]. At the same time, new trends in scholarly literature are developing an argument around responsible data management, ethics issues, and artificial intelligence-driven governance. Such an interdisciplinary trend demonstrates the complexity of data stewardship research, which encompasses technological, managerial, and institutional perspectives.

Even though the importance of data stewardship as an organizational competence has been increasingly realized, there is still a lack of integration of literature related to this issue. The different definitions, theories, and methods applied to the study of data stewardship practice contribute to the lack of clarity concerning how this research area has evolved, what themes are currently prevailing, and what new themes should be explored. It is vital to conduct an overview of research trends in order to understand the intellectual structure of data stewardship research, the changes that have taken place in terms of the development of concepts, and where future research may go. The technique of research mapping, including bibliometric analysis, offers many advantages for a systematic analysis of publications, key contributions, research themes, and research areas in a particular field [11].

While the concept of data stewardship has been attracting growing research interest among scholars, there are still a number of issues that need to be resolved. First of all, data stewardship is understood differently in different pieces of literature leading to inconsistency in the perception of its purposes and functions as well as its influence on organizations.

Secondly, the current scientific literature is spread in various disciplines and, thus, makes it hard to trace the general evolution and knowledge base of this field. Thirdly, only little research exists concerning the evolution of themes and publications in data stewardship literature as well as collaborative trends and future prospects in this field.

The current study is designed to investigate the trends in research on data stewardship in the organizational environment through the study of its patterns of development, intellectual structure, dominant topics, and future directions. In particular, the research is intended to explore the trends in publications, identify key scholars and papers, trace the changes in the themes of the field, and give insights into future avenues of research in data stewardship. Through systematic analysis of existing literature, this research will contribute to the better understanding of the development of data stewardship as an organizational skill.

2. METHODS

This research uses a bibliometric analysis framework for studying the evolution and research trends in data stewardship within organizations. The bibliometric analysis framework is a quantitative framework through which researchers are able to evaluate the scientific literature by analyzing the patterns of publication, the significant publications, the structure of knowledge and the research themes in the area [11]. This framework has been chosen because the aim of this research is not only to conduct a literature review but to provide an overview of how research in this area has evolved over time. The analysis will be conducted on scientific documents concerning data stewardship, especially within the context of organizations, data

governance, data management, information systems and organizational capabilities.

Data gathering was performed using the Scopus database as the main source of scientific publications due to the wide coverage of peer-reviewed journals and international studies. The search was carried out by means of keywords related to the topic of the research. These keywords were "data stewardship," "data steward," "data governance," and "organizational data management." The research covered only those publications that contained the keywords in their titles, abstracts, or keywords in order to be relevant to the topic of the study. All found papers were analyzed according to specific criteria of inclusion and exclusion, such as relevance, type of publication, and scope of the study. Only peer-reviewed journal papers and conference proceedings in English were selected for analysis.

Bibliometric analysis and visualization approaches like VOSviewer and bibliometric software programs were used in the analysis of the gathered data. Performance analysis was performed to measure publication trends, citations, and significant contributions, along with the science mapping analysis for relationships among authors, keywords, and themes of research. Co-occurrence of keywords analysis was done to determine significant topics and theme clusters, and trend analysis was performed to find out the progression of interest areas over time. By the processes mentioned above, this study is able to present an organized view of the current state of data stewardship research, as well as suggest possible directions for future studies in organizational data management and governance.

3. RESULTS AND DISCUSSION

3.1 Co-Authorship Analysis

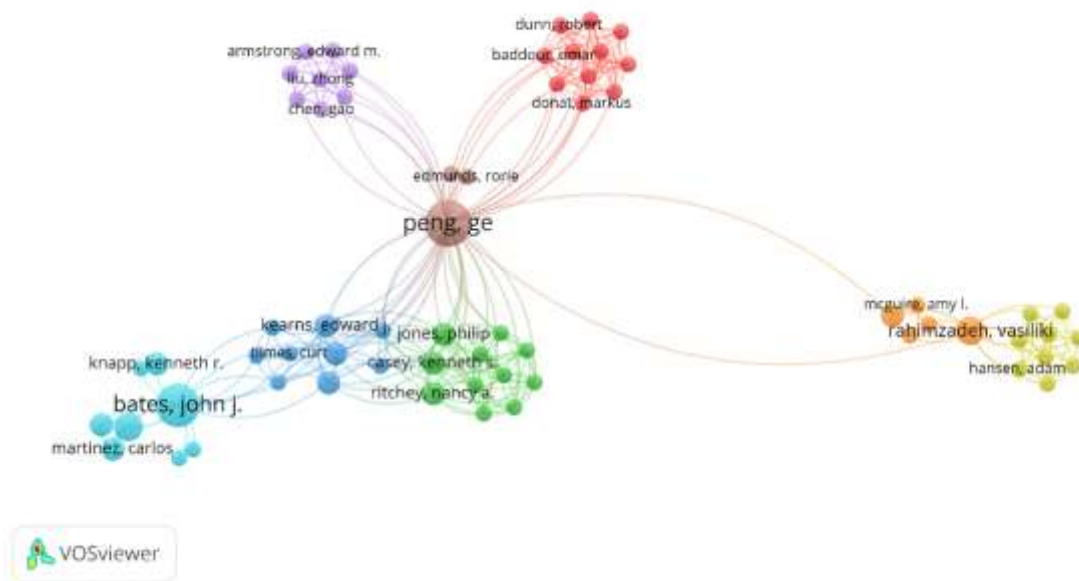


Figure 1. Author-Level Visualization

Source: Data Analysis

The network analysis shown in Figure 1 presents the intellectual structure of research on data stewardship in organizational settings, showing several clusters of significant scholars that are connected to each other. The largest and the most central node belongs to Peng, Ge, which signifies great influence and connectivity in this field of study. This scholar plays a mediating role in linking many clusters of knowledge, especially for studies concerning information management, health care data, and organizational data governance. Other significant scholars such as Bates, John J., Dunn, Robert, and Rahimzadeh, Yasik are associated with different research clusters contributing in their own way to the perspective of data stewardship.

The clusters colored blue and green can be understood as comprising mainly those studies that are based on issues of data management and information systems with emphasis on data quality, information governance, and responsible data handling. The studies represented by the red cluster are those that have been done in relation to healthcare information and data governance applications. The yellow cluster shows new linkages between data stewardship and general issues of organization or technology management. Some linkages that exist among the peripheral clusters in smaller numbers show that the field of data stewardship is still fragmented and provides scope for future research.

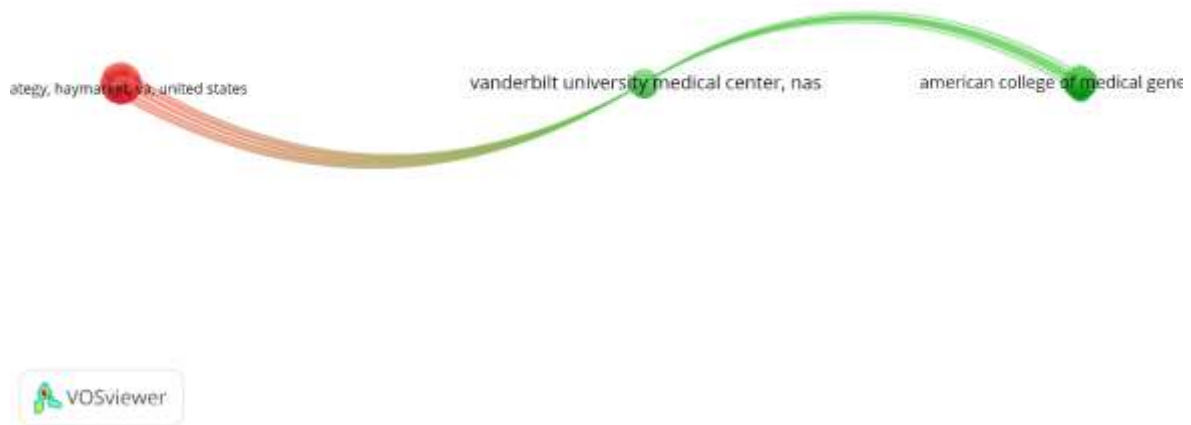


Figure 2. Institution-Level Visualization

Source: Data Analysis

Figure 2 shows a relatively limited but meaningful connection structure among organizations involved in data stewardship research. The network consists of three main institutions, with Vanderbilt University Medical Center, NAS positioned as the central connecting node between Strategy, Haymarket, VA, United States and the American College of Medical Genetics. This structure indicates that research collaboration in this field is strongly influenced by healthcare and biomedical institutions, where data stewardship practices are closely associated with clinical data management, genomic information governance, and

institutional data-sharing initiatives. The strong connecting links suggest that Vanderbilt University Medical Center plays an important bridging role in facilitating knowledge exchange between different research communities. The limited number of connected institutions also indicates that organizational data stewardship research remains concentrated within specific domains, particularly healthcare, creating opportunities for broader collaboration with industries, technology organizations, and business management researchers.

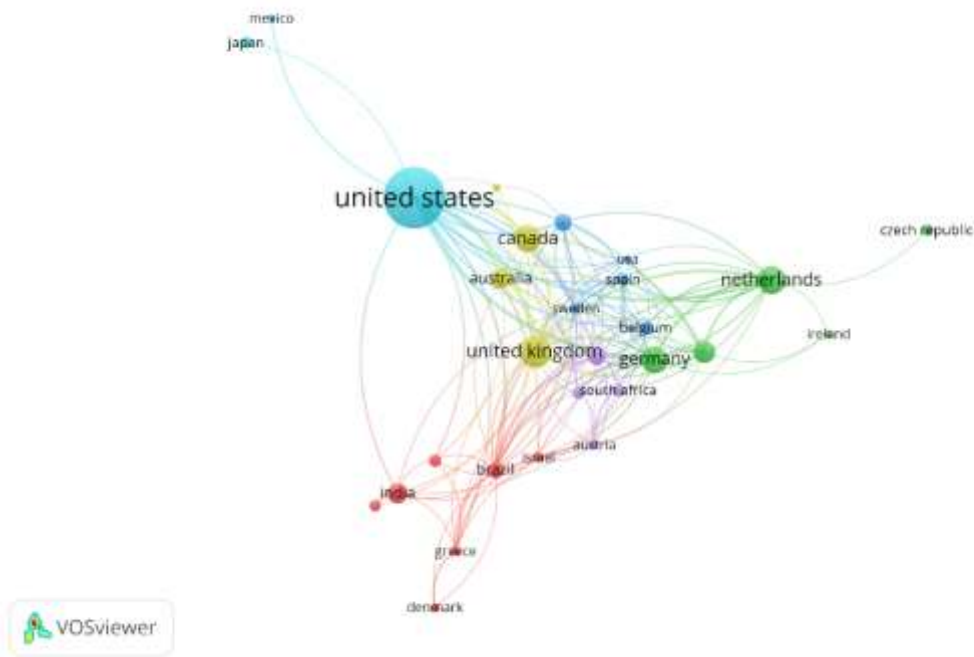


Figure 3. Country-Level Visualization
Source: Data Analysis

Figure 3 highlights the distribution of global data stewardship research and cooperation on an international level. The United States is shown to be the most prominent and central country because of its large node size and many connections to other countries. This shows that the United States has a dominant position in contributing to the growth of data stewardship knowledge by conducting research and collaborating with other countries. Some other prominent countries like the United Kingdom, Netherlands, Germany, Canada, Australia, and Sweden have a close network of research activities, implying cooperation among the developed countries, specifically in the areas of information governance and digital transformation. The network further

highlights the existence of regional clusters that show the research contributions made by each region. In the European cluster, countries such as the Netherlands, Germany, Belgium, and Ireland show strong collaboration within the region, whereas countries such as Brazil, India, South Africa, and Israel show emerging contribution through their involvement in global research on data stewardship. The existence of ties between developed and developing countries shows the existence of international knowledge sharing, but the small size of the nodes for many countries suggests the dominance of research efforts in just a few countries. The development of future research can be improved through greater involvement of other regions.

3.2 Citation Analysis

Table 1. The Most Impactful Literatures

Citations	Authors and year	Title
14667	[12]	Comment: The FAIR Guiding Principles for scientific data management and stewardship
239	[13]	The FAIR guiding principles for data stewardship: Fair enough?
173	[14]	Data governance and stewardship: Designing data stewardship entities and advancing data access

Citations	Authors and year	Title
121	[15]	Scientific data stewardship of international satellite cloud climatology project B1 global geostationary observations
117	[16]	A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth
100	[17]	ENSURING THE INTEGRITY, ACCESSIBILITY, AND STEWARDSHIP OF RESEARCH DATA IN THE DIGITAL AGE
93	[18]	Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 health policy meeting
80	[19]	Erratum: Addendum: The FAIR Guiding Principles for scientific data management and stewardship (Scientific data (2016) 3 (160018))
62	[20]	Metadata stewardship in nanosafety research: Community-driven organisation of metadata schemas to support fair nanoscience data
55	[21]	Data-driven government: Cross-case comparison of data stewardship in data ecosystems

Source: Scopus, 2026

Table 1 shows the major literature in the field of data stewardship research, which includes the top works which contributed to the development of this sphere of research in the most significant way. The work with the highest citation rate is [12] entitled "The FAIR Guiding Principles for Scientific Data Management and Stewardship" with 14,667 citations. The work plays an essential part in setting up principles for the making research data Findable, Accessible, Interoperable, and Reusable (FAIR). This paper is one of the main references on data stewardship which offers a worldwide approach to improving data management practices and scientific data sharing.

The other significant publications further show that data stewardship is indeed multidisciplinary and covers disciplines like environmental sciences, health care, research integrity, and digital government. The articles from [15] and [16] focus on scientific and environmental data stewardship whereas [18] discusses issues in data governance and stewardship in health sector. Recent studies like those of [20] and [21] suggest that there is a growing trend towards metadata stewardship and data ecosystems among others.

3.3 Keyword Analysis

Co-Occurrence

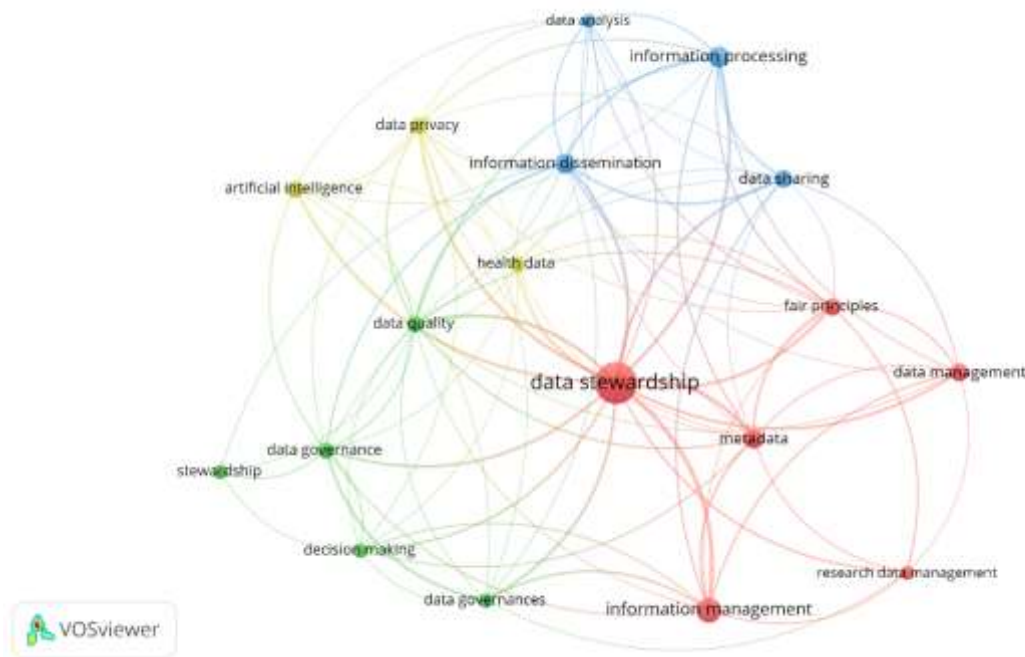


Figure 4. Network Visualization

Source: Data Analysis

Figure 4 presents the conceptual framework as well as the main research areas of data stewardship research. It can be observed from the figure that the keyword "data stewardship" is the central and most influential keyword of all the other keywords which are connected with different research areas. Being located in the center of the map means that "data stewardship" serves as a connecting concept for all these dimensions of data management. The other keywords presented in Figure 4 also prove that the scope of data stewardship research has broadened from pure technical aspects of data handling to other dimensions as well.

The red cluster is associated with the basic cluster of data management and FAIR principles, which includes data management, metadata, research data management, information management, and FAIR principles as keywords. It indicates that in the first phase of the study of data stewardship, much attention was paid to data organization, preservation, documentation, and distribution. The relevance of the relation between the data stewardship and the FAIR principles can be explained by the necessity to ensure findability, accessibility,

interoperability, and reusability of organizational and scientific data. Metadata is another key element, as it should be properly documented for the sake of stewardship.

The green cluster highlights the link between data governance, data quality, and decision-making within organizations. Some important terms like data governance, data quality, decision-making, and stewardship show that there is an increasing research trend towards looking at data from an organizational asset perspective. The green cluster signifies the fact that good data stewardship necessitates the setting of clear roles, data governance structure, and methods to guarantee data quality for decision-making purposes. The association between data quality and data governance shows the increasing trend of considering stewardship as a capacity in organizations.

The blue and yellow clouds show future trends in data stewardship. These include data sharing, health data, artificial intelligence, data privacy, and information processing. The inclusion of artificial intelligence and data privacy implies that current studies are moving toward the adoption of technology in a responsible

manner as well as the ethical use of data. In addition, health data is also present as an application area for data stewardship.

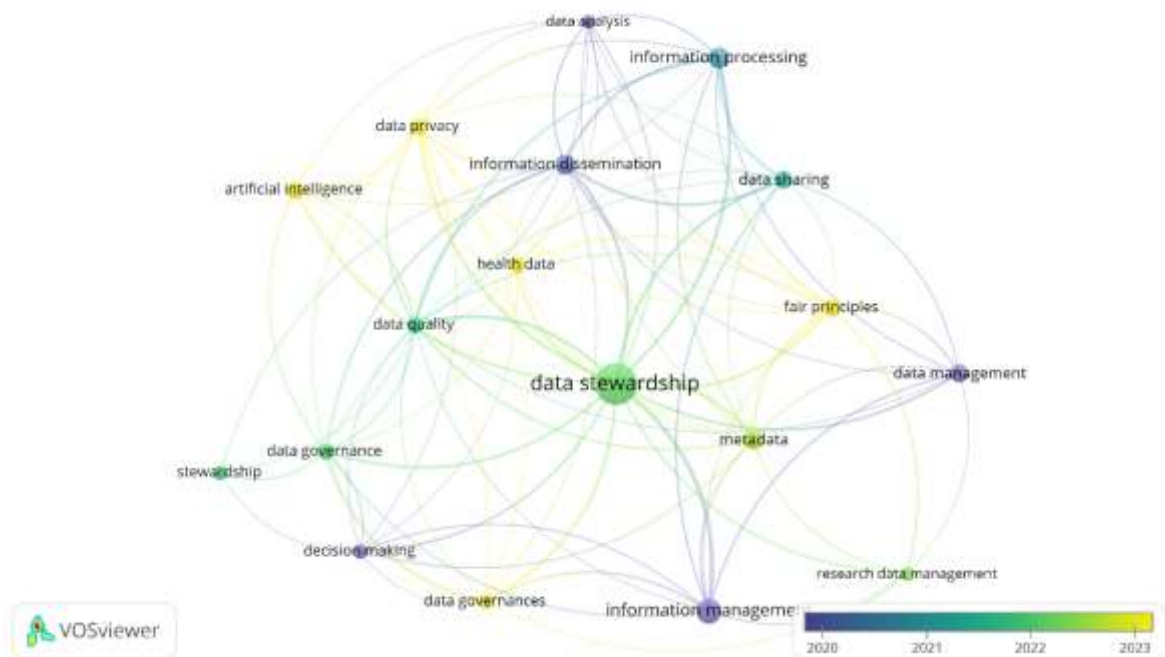


Figure 5. Overlay Visualization

Source: Data Analysis

Figure 5 shows the evolution in time of keywords related to the field of data stewardship. The color coding of the graph reflects the average publication period of the corresponding keywords, whereby the darker blue shade is for older research topics while the lighter yellow shade reflects newer ones. As can be seen, the central node of the term “data stewardship” is the most connected one with an average publication period of around mid-time of the period under observation. This confirms its continued relevance and function as an intermediary connecting different areas of research.

The research themes of earlier periods, depicted in darker blue and purple nodes, are more relevant to information management, data analysis, data management, decision making, and information dissemination. The themes above are indicative of the early research in data stewardship, which was focused on the technical and operational issues of data management within organizations and

sciences. The research in this period was aimed at enhancing the organization and processing of data as well as the management of data research resources. The close association of data stewardship with information management suggests its origin from data management disciplines.

On the other hand, new research themes, which appear in the form of green to yellow colored nodes, emphasize the newly raised issues in relation to data privacy, artificial intelligence, FAIR principles, data governance, and responsible data sharing. Such new trends in data stewardship research show the transition of this research towards the strategic and ethical aspects of data stewardship where organizations need to ensure data security, transparency, interoperability, and responsible usage. The inclusion of artificial intelligence and privacy as new themes in the research demonstrates increasing concern for managing the risks associated with data in the advanced digital world.

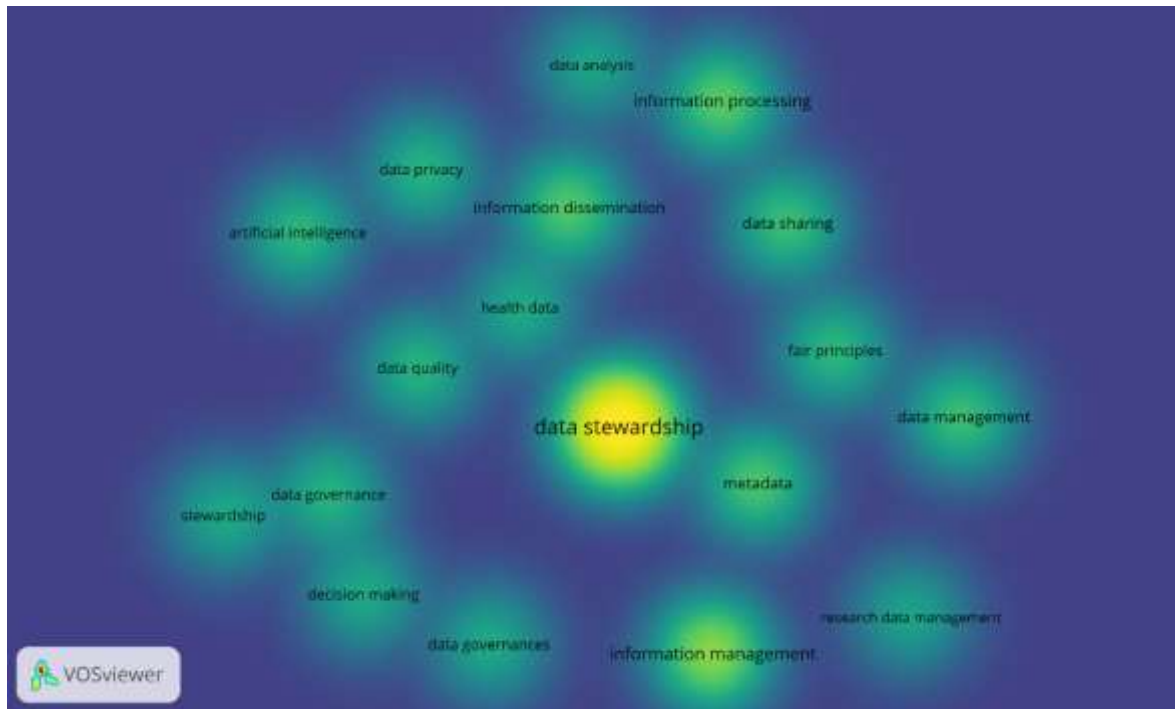


Figure 6. Density Visualization

Source: Data Analysis

Figure 6 shows the distribution and relevance of keywords in research regarding data stewardship. The area with the greatest density of keywords is shown at the center of the keyword "data stewardship", which indicates the presence of the brightest color region; this proves the predominance and frequent occurrence of the concept in the literature. This central placement proves that data stewardship is the primary theme of research, which is used to tie other related themes such as data management, governance, quality, sharing, and information management together.

Some of the additional high-density clusters have been observed in relation to terms like data management, information management, metadata, data governance, and data quality, and these clearly show that these terms are forming the basic body of knowledge for data stewardship research. This clearly shows the need for the development of some formalized process for managing data and making sure that it is maintained and documented and its quality is assured.

The keywords with lower density, but which are strategically located, such as

artificial intelligence, data privacy, FAIR principles, data sharing, and health data, reflect growing areas of research interest in the field. Though the keywords seem less dense in terms of frequency, it indicates growing concern about integrating the issues related to data stewardship with new technology and ethics. Artificial intelligence and data privacy indicate the problem of proper use of data.

4. CONCLUSION

This paper suggests that research on data stewardship in organizations has evolved into a multidisciplinary area combining data management, governance, quality assurance, and proper data use. Based on the results of bibliometric analysis, it is found that the topic of data stewardship has shifted from the first interest related to scientific data management and information organization to more strategic issues connected with data governance, FAIR principles, data privacy, artificial intelligence, and data-driven decision-making. It is identified that data stewardship, data management, data governance, metadata, and data quality are the key themes forming the

intellectual structure of the discipline, and the most significant publications, such as FAIR Guiding Principles, served as the background for the development of responsible data practices. In addition, the analysis of the co-authorship network shows that the activities of researchers continue to be geographically clustered, and research is conducted in the US and healthcare organizations specifically. However, international collaboration is still growing. Thematic evolution suggests the

shift from operational data management to the consideration of data stewardship as an organizational competency ensuring transparency, responsibility, innovation, and sustainability of data ecosystems. The next step for future research would be exploring data stewardship practice in different organizational contexts, including the business and public sectors, as well as new digital environments.

REFERENCES

- [1] A. Nag and S. Mishra, "Revitalizing Mining Heritage Tourism: A Machine Learning Approach to Tourism Management," *J. Min. Environ.*, vol. 15, no. 4, pp. 1193–1225, 2024, doi: 10.22044/jme.2024.13770.2554.
- [2] M. Sigala, "Implementing social customer relationship management: A process framework and implications in tourism and hospitality," *Int. J. Contemp. Hosp. Manag.*, vol. 30, no. 7, pp. 2698–2726, 2018, doi: 10.1108/IJCHM-10-2015-0536.
- [3] P. Gupta and A. Gupta, "Strategic Integration of Sustainability in Talent Acquisition and Employer Branding: Advancing Competitive Edge and Social Responsibility," *Managing for Sustainability: The Role Of Human Resource Management In Advancing Sdg Goals*. World Scientific Publishing Co., JECRC University, Jaipur, India, pp. 93–114, 2025. doi: 10.1142/9789819807215_0005.
- [4] H. R. Gilani, J. L. Innes, and A. De Grave, "The effects of seasonal business diversification of British Columbia ski resorts on forest management," *J. Outdoor Recreat. Tour.*, vol. 23, pp. 51–58, 2018, doi: 10.1016/j.jort.2018.07.005.
- [5] C. G. Di Camillo *et al.*, "Towards blue diving: analysis and solutions to prevent scuba diving impact in the Mediterranean sea," *Ocean Coast. Manag.*, vol. 269, 2025, doi: 10.1016/j.ocecoaman.2025.107833.
- [6] E. Alanzi, M. Sarabdeen, H. Z. Almugren, and A. C. M. Kijas, "Regional Tourism Development: The Role of Sustainable Practices, Logistics Infrastructure, Uncertainty, Safety and Economic Environment of the Countries in Attracting Inbound Tourists," *Sustain.*, vol. 18, no. 8, 2026, doi: 10.3390/su18083968.
- [7] A. Singh, "Mapping Trends, Drivers, and Barriers in AI-Driven Medical Tourism," *Transforming Global Healthcare with AI-Driven Medical Tourism to Advance SDG 3*. Nova Science Publishers, Inc., Maharishi University of Information Technology, Uttar Pradesh, Noida, India, pp. 237–256, 2026. [Online]. Available: <https://www.scopus.com/pages/publications/105042928603?origin=resultslist>
- [8] S. Al-Tahat, S. Bani-Khaled, Z. Jaradat, M. Mansour, and A. M. Al-zoubi, "State ownership as a moderator in the relationship between board characteristics and ESG performance: evidence from Asia-Pacific markets," *J. Bus. Socio-economic Dev.*, pp. 1–22, 2025, doi: 10.1108/JBSED-05-2025-0145.
- [9] A. Erba, F. C. Franzeck, V. Hinic, A. Egli, and M. Osthoff, "Utilization of a Meningitis/Encephalitis PCR panel at the University Hospital Basel – a retrospective study to develop a diagnostic decision rule," *Front. Med.*, vol. 11, 2024, doi: 10.3389/fmed.2024.1351903.
- [10] R. T. Sheffield, "More than acid-free folders: Extending the concept of preservation to include the stewardship of unexplored histories," *Libr. Trends*, vol. 64, no. 3, pp. 572–584, 2016, doi: 10.1353/lib.2016.0001.
- [11] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021.
- [12] M. D. Wilkinson *et al.*, "Comment: The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, 2016, doi: 10.1038/sdata.2016.18.
- [13] M. Boeckhout, G. A. Zielhuis, and A. L. Bredenoord, "The FAIR guiding principles for data stewardship: Fair enough?," *Eur. J. Hum. Genet.*, vol. 26, no. 7, pp. 931–936, 2018, doi: 10.1038/s41431-018-0160-0.
- [14] S. Rosenbaum, "Data governance and stewardship: Designing data stewardship entities and advancing data access," *Health Serv. Res.*, vol. 45, no. 5 PART 2, pp. 1442–1455, 2010, doi: 10.1111/j.1475-6773.2010.01140.x.
- [15] K. R. Knapp, "Scientific data stewardship of international satellite cloud climatology project B1 global geostationary observations," *J. Appl. Remote Sens.*, vol. 2, no. 1, 2008, doi: 10.1117/1.3043461.

- [16] P. Kansakar and F. Hossain, "A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth," *Space Policy*, vol. 36, pp. 46–54, 2016, doi: 10.1016/j.spacepol.2016.05.005.
- [17] D. Kleppner *et al.*, *ENSURING THE INTEGRITY, ACCESSIBILITY, AND STEWARDSHIP OF RESEARCH DATA IN THE DIGITAL AGE*. Massachusetts Institute of Technology, Cambridge, United States: National Academies Press, 2009. doi: 10.17226/12615.
- [18] G. Hripcsak *et al.*, "Health data use, stewardship, and governance: Ongoing gaps and challenges: A report from AMIA's 2012 health policy meeting," *J. Am. Med. Informatics Assoc.*, vol. 21, no. 2, pp. 204–211, 2014, doi: 10.1136/amiajnl-2013-002117.
- [19] M. D. Wilkinson *et al.*, "Erratum: Addendum: The FAIR Guiding Principles for scientific data management and stewardship (Scientific data (2016) 3 (160018))," *Sci. data*, vol. 6, no. 1, p. 6, 2019, doi: 10.1038/s41597-019-0009-6.
- [20] A. G. Papadimitriou *et al.*, "Metadata stewardship in nanosafety research: Community-driven organisation of metadata schemas to support fair nanoscience data," *Nanomaterials*, vol. 10, no. 10, pp. 1–49, 2020, doi: 10.3390/nano10102033.
- [21] W. van Donge, N. Bharosa, and M. F. W. H. A. Janssen, "Data-driven government: Cross-case comparison of data stewardship in data ecosystems," *Gov. Inf. Q.*, vol. 39, no. 2, 2022, doi: 10.1016/j.giq.2021.101642.