

RESEARCH ARTICLE

Mapping the evolution of data governance scientific research

Hossein Hassani¹ , Xu Huang² and Steve MacFeely³

¹International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria

²Headington Campus, Oxford Brookes Business School, Oxford Brookes University, Oxford, OX3 0BP, UK

³Statistics and Data Directorate, OECD, Paris, France

Corresponding author: Hossein Hassani; Email: hassani.stat@gmail.com

Received: 13 November 2024; **Revised:** 28 April 2025; **Accepted:** 14 May 2025

Keywords: academic literature; data governance; research; theme; trend

Abstract

Data governance has emerged as a pivotal area of study over the past decade, yet despite its growing importance, a comprehensive analysis of the academic literature on this subject remains notably absent. This paper addresses this gap by presenting a systematic review of all academic publications on data governance from 2007 to 2024. By synthesizing insights from more than 3500 documents authored by more than 9000 researchers across various sources, this study offers a broad yet detailed perspective on the evolution of data governance research.

Policy Significance Statement

Data governance is a cornerstone of effective decision-making in the digital age, ensuring the ethical, secure, and efficient management of data to support societal, organizational, and individual interests. Despite its critical importance, a comprehensive understanding of the academic discourse surrounding data governance has been lacking. This policy emphasizes the need for robust, interdisciplinary research to inform evidence-based governance frameworks that address emerging challenges, such as data privacy, security, and equity. The findings from this systematic review highlight the evolution of data governance research and its implications for public policy, corporate responsibility, and global collaboration. Policymakers are urged to adopt principles rooted in transparency, accountability, and inclusivity, leveraging the collective knowledge presented in this analysis to design adaptive and sustainable data governance strategies that align with societal values and technological advancements.

1. Introduction

Over the last decade, data governance has solidified its role as a crucial area of academic inquiry, reflecting the growing complexity and importance of managing data in various sectors and policy interest and real-world applications. The significance of data governance in official statistics has also become increasingly clear in recent years, as organizations within national statistical systems acknowledge its essential role in promoting effective and responsible data management (see, for example, United Nations Statistical Commission, 2013; International Monetary Fund [IMF], 2014; United Nations, 2017; Eurostat, 2020; World Bank, 2020) and as National Statistical Offices (NSO) come to terms with integrating data from a wider variety of data sources.

Key contributors to the data governance landscape include international organizations such as the *United Nations (UN)*, *UN System Chief Executives Board for Coordination (UN CEB)*, *UN HLCP*, *World Bank*, *WHO*, *OECD*, *G7*, *G20*, *G77*, *African Union*, *UNCTAD*, *CCSA*, *European Union*, *Amnesty International*, and Google, among others (for further details, see, for example, G20, 2019; MacFeely, 2020; MacFeely et al., 2020; G7. Carbis Bay G7 Summit Communiqué, 2021; G7. G7 Roadmap for Cooperation on Data Free Flow with Trust, 2021; Me et al., 2021; United Nations, 2021; World Bank, 2021; European Union, 2022; MacFeely et al., 2022; G77, 2023; Hassani and MacFeely, 2023; United Nations, 2023; UNCTAD, 2021).

Despite the surge in interest and research, a comprehensive overview of the academic literature in this field has been lacking. This paper aims to bridge this gap by conducting an extensive review of data governance publications from 2007 to 2024, covering over 3500 documents authored by more than 9000 researchers across numerous sources. In order to capture the most comprehensive picture of related publications, the only selection criterion is containing the term “data governance” in the title and abstract. This research aims to extensively review all available data-governance-related publications that we could find from the Dimensions database when this research was conducted in 2024. The earliest publication we could find is from 2007.

This review captures a period of remarkable growth in data governance research, with an impressive annual publication growth rate of 43.4%. The high rate of international collaboration—reflected by the 12.7% of papers involving co-authors from different countries—underscores the global relevance of data governance. The field’s influence is further emphasized by an average citation rate of 7 per document, while the substantial number of references (42,396) highlights the thorough engagement and cross-disciplinary interest in this topic.

By analyzing publication trends, authorship patterns, and citation impact, this paper provides a foundational overview of the data governance research landscape. As the first systematic synthesis of its kind, this study offers valuable insights for researchers, practitioners, and policymakers who seek to understand the trajectory and impact of data governance, setting a benchmark for future research in this essential field.

To provide a structured exploration of the data governance research landscape, this paper is organized as follows: Section 2 examines the sources and global distribution of scientific production; Section 3 analyzes publication trends and collaborative networks; Section 4 explores thematic evolution over time; Section 5 highlights relevant keywords and affiliations within the literature; and Section 6 categorizes research according to major fields and alignment with the UN Sustainable Development Goals (UN SDGs). Together, these sections offer a comprehensive overview for researchers, policymakers, and practitioners, setting a foundation for future inquiry in this essential field.

2. Data source

This research analyzes a total of 3543 documents gleaned from 1785 distinct sources, covering the period from January 2007 to July 2024, which represents the latest data available. The only selection criterion is containing the term “data governance” in the title and abstract, and starting from the first available publication in 2007, we aim to capture all data available from the Dimensions database at the time this research was conducted. These documents have been authored by a total of 9366 researchers, of which 787 of these being single-authored works. Each document comprised an average of approximately 3.4 co-authors, highlighting the collaborative nature of research in this field. Furthermore, the documents collectively reference 42,396 sources, indicating a robust foundation of literature and knowledge in data governance.

Each document within this dataset has garnered an average of seven citations, suggesting that data governance is a highly relevant and actively researched topic among scientists. This level of citation reflects the growing interest and importance of data governance in various academic and professional domains.

Country Scientific Production

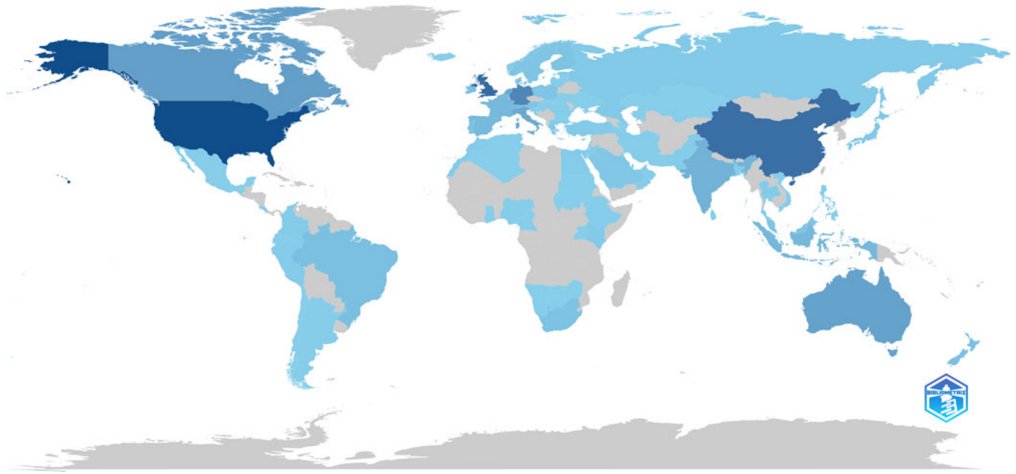


Figure 1. Global distribution of scientific publications by country (2007–2024). Data source: dimensions database.

2.1. Global distribution of scientific production

Figure 1 illustrates the global distribution of scientific production by country, highlighting the contributions to scientific research output in the data governance field on a world scale. Each country is shaded according to its relative level of scientific output, with darker shades representing higher levels of production.

Key observations:

1. **High Output Countries:** The United States (15%), UK (12%), China (8%), and Germany (6%) are prominently leading research in this field as the top four countries, with darker shades indicating substantial contributions to global scientific research in the data governance field. It is worth noting that the leading countries are in line with their overall scientific success across all fields, e.g., the US (20%), China (14%), UK (6%), and Germany (5%) are also the top four countries according to the Scimago Journal & Country Rank (SJR) country ranking for all subject areas based on their record of documents from 1996 to 2024 (SJR. Scimago Journal & Country Rank, 2024).
2. **Moderate Output Regions:** Countries in South America, parts of Asia, and Oceania show moderate levels of scientific production, as reflected by lighter shades of blue.
3. **Low Output and Data-Deficient Areas:** Certain regions, particularly in Africa and some parts of the Middle East, have minimal representation in scientific output, highlighted by lighter or grayed-out areas. This may reflect lower research funding, infrastructure, or data unavailability.

This map underscores the uneven distribution of scientific research across the globe in the data governance field, with significant contributions from advanced economies and growing outputs from emerging nations. For instance, Indonesia has the 10th highest output for data governance field globally, while it is ranked 28th for all subject areas in general in the SJR country rank (SJR. Scimago Journal & Country Rank, 2024); similarly, Ireland is ranked 16th for data governance field scientific productions, while it has a place in 43rd for the SJR all subject area ranking. It is great to see the growing outputs from these countries, which are playing a more and more important role in data and technology research advancements and innovation and breaking the dominance of advanced economies. This map highlights the need for collaboration, funding, and capacity building in underrepresented regions to promote a more balanced global research landscape.

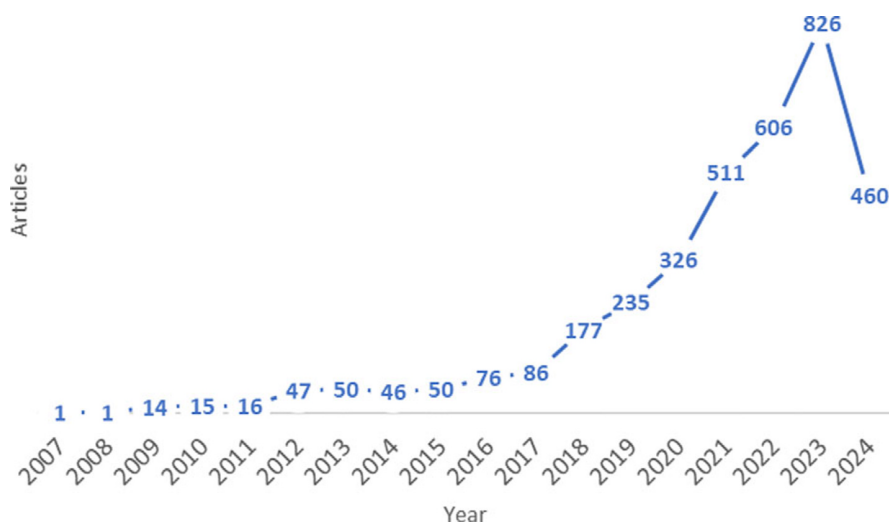


Figure 2. Annual growth in scientific publications on data governance (2007–2024). Data source: Dimensions database.

2.2. Trends in annual scientific production

Figure 2 illustrates the growth in annual scientific production from 2007 to 2024, as measured by the number of published articles.

Key observations:

1. **Steady Growth (2007–2018):** From 2007 to 2018, scientific production grew at a steady but gradual rate, indicating a period of consistent growth in research output.
2. **Significant Increase (2019–2023):** Starting around 2019, there was a marked acceleration in scientific publications, peaking in 2023. This increase may be attributed to a combination of factors, including enhanced research funding, technological advancements, and potentially a response to global events, such as the COVID-19 pandemic, which spurred scientific exploration across various disciplines.
3. **2024 Mid-Year Data:** The data for 2024 shows a partial count up to July, slightly below the 2023 peak. However, if the trend continues as expected through the latter half of the year, 2024 is likely to match or exceed the high output observed in 2023.

The exponential rise in scientific production over recent years underscores a growing global emphasis on research and innovation. Despite a slight drop in mid-2024, the trend suggests that the field's momentum is likely to carry forward, making 2024 another potentially high-output year for scientific publications.

2.3. International collaboration in scientific research

Figure 3 visualizes the collaboration patterns between countries in scientific research. Each node represents a country, and the size of the node indicates the prominence of that country in international collaborations, with larger nodes representing more central players with more scientific productions. The connections (edges) between nodes signify collaborative links, and the color coding groups countries into different collaboration clusters. Please note that the color coding is automatically assigned based on collaborative connections among countries (nodes), not geographical locations. It is also acknowledged by the authors that there are nodes which did not have their country name displayed. This is due to the display of nodes, and their country names are prioritized for more central players, following the country-level production ranking. Differences in the amount of production following the ranking list (more central

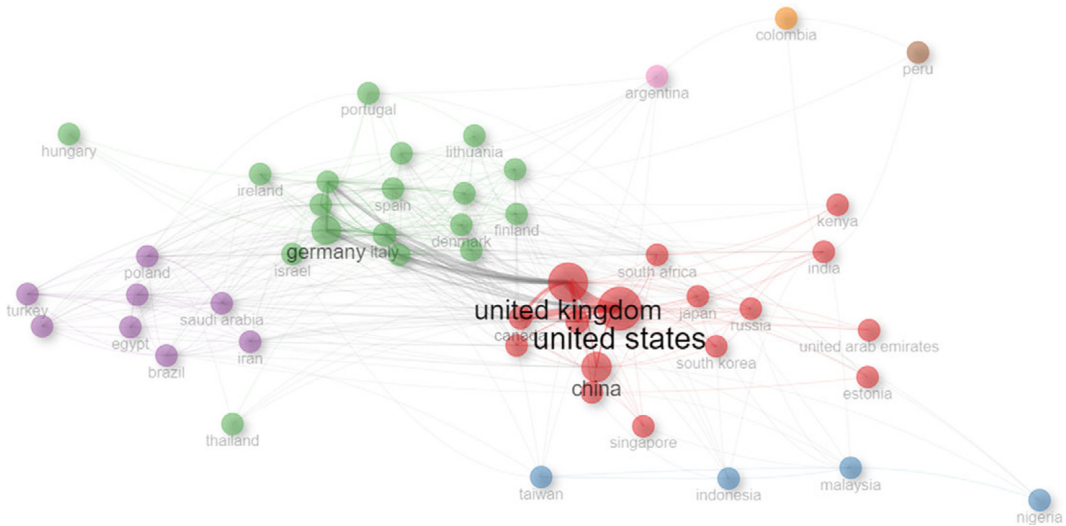


Figure 3. International collaboration network in data governance research. Nodes (countries) of different sizes (according to scientific production country ranking, where a larger node indicates a more central, productive player) are grouped and connected based on research collaborations.

players and less central players) may not be visually apparent in the figure below via the size of those nodes. Increasing the number of named nodes in the data visualization setting will result in some unnamed nodes for the same reason, given the differences after the top 10 countries get smaller. Hence, this current figure focuses on the top 30 countries and their collaborative connections to provide the best balance in terms of clarity, quantity, and quality.

Key observations:

1. **Prominent Collaborators:** The United States, the United Kingdom, and China stand out as central hubs, with the largest nodes and extensive connections. These countries have robust international partnerships, spanning both developed and emerging research nations, which positions them as influential players in the global scientific landscape.
2. **Regional Clusters:**
 - **Green Cluster:** Primarily consists of European countries like Germany, Italy, Spain, and Portugal, showing strong intra-European collaborations. This cluster highlights the collaborative framework within the European Union, encouraging joint research across borders and also General Data Protection Regulation (GDPR)-related activities.
 - **Red Cluster:** Includes major research players from Asia and other global regions, such as China, Japan, South Korea, and India, connecting strongly with the United States and the United Kingdom.
 - **Purple Cluster:** Includes countries like Turkey, Saudi Arabia, Brazil, and Egypt, representing collaborative ties with both European and Asian clusters, suggesting emerging partnerships.
3. **Emerging and Peripheral Collaborators:** Countries like Nigeria, Indonesia, and Malaysia (in the blue cluster) have smaller nodes, indicating growing but relatively limited global research partnerships. However, their connections within their respective clusters show emerging collaboration potential.

This network highlights the structure of global scientific collaboration, emphasizing the central role of certain high-output countries and the regional clustering of collaboration patterns. It also suggests potential areas for fostering increased global cooperation, particularly among emerging economies and

underrepresented regions. Strengthening these international partnerships could contribute to a more balanced and inclusive global research ecosystem.

3. Thematic evolution in data governance research (2007–2024)

The thematic evolution of data governance research, visualized in Figure 4, reveals significant shifts in focal topics across four distinct periods: 2007–2017, 2018–2020, 2021–2023, and 2024. This progression underscores the field’s adaptive nature, as research priorities shift to address emerging challenges and technologies. The thematic analysis was conducted based on co-word analysis and clustering using the Bibliometrix package in R (Aria and Cuccurullo, 2017; Aria et al., 2022). In brief, the keywords and abstract of all the collected documents are used as the textual body, which is preprocessed so that all the unique words used are identified to form a list of vocabulary. Each text can be considered as a vector, and the co-occurrence matrix is built from the set of vectors, from which the association strength is evaluated, and a community detection procedure (similar to clustering) is then conducted. This will then allow detected communities to be associated with various topics, so that the thematic evolution can be plotted accordingly. More detailed steps and algorithms behind the scenes can be found in the Bibliometrix R package manual (Aria and Cuccurullo, 2017; Aria et al., 2022).

2007–2017: Foundational themes

In the early years, data governance research focused on foundational concepts, with recurring themes such as **health**, **data**, and **privacy**. Other prominent topics included **collaboration**, **compliance**, and **service**

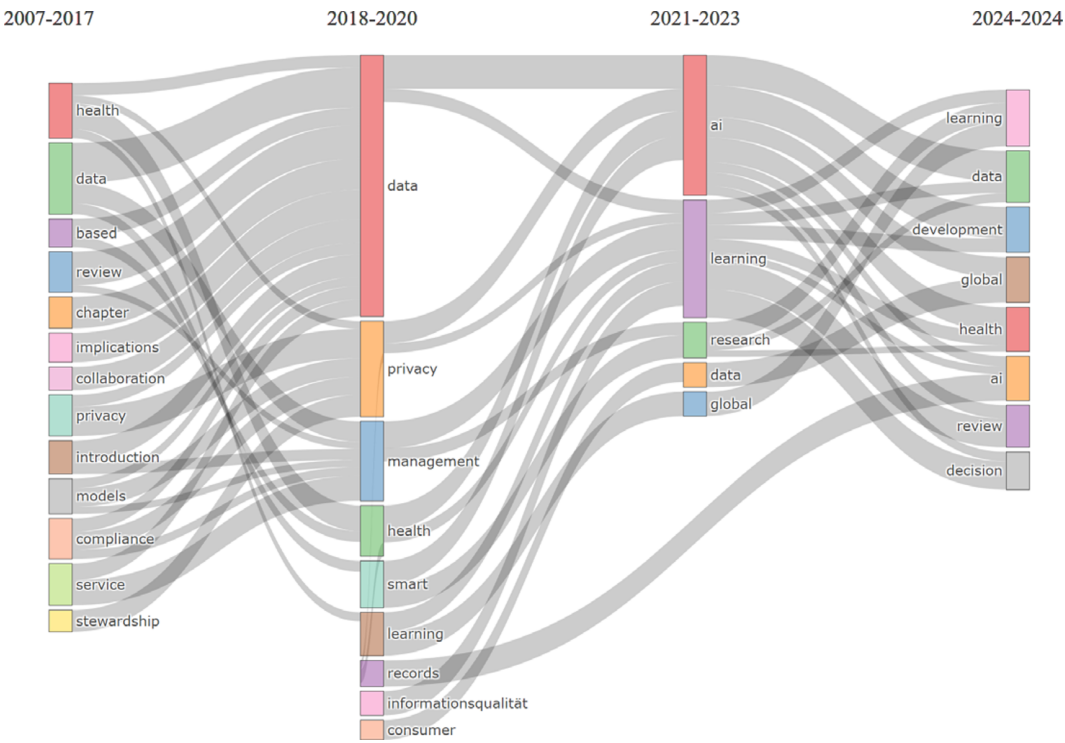


Figure 4. Thematic evolution of data governance research (2007–2024). Data source: Dimensions. Topics of each period (in colored blocks, block sizes reflect higher focus and occurrence) are connected, reflecting research priority shifts.

stewardship, reflecting the early need to establish frameworks for data management and data quality. This period laid the groundwork for defining data governance, emphasizing regulatory compliance and basic stewardship practices.

2018–2020: Expansion and privacy concerns

The period from 2018 to 2020 saw a narrowing of focus, with **data** and **privacy** emerging as dominant themes. This shift aligns with the introduction of stringent data privacy regulations worldwide, such as the EU's GDPR, which sparked increased academic and practical interest in data privacy as a crucial aspect of governance. Topics like **management**, **health**, and **smart systems** began to intersect with privacy, indicating the need for governance structures that balance data utility and privacy across diverse applications.

2021–2023: AI and learning systems

As the field progressed into 2021–2023, artificial intelligence (**AI**) and **learning** became central themes, reflecting the growing integration of AI and machine learning into data governance practices. This period saw an increased focus on **research** and **global** data governance, as researchers explored frameworks for handling the complexities introduced by AI. With data increasingly used to train machine learning models, issues surrounding data bias, accountability, and transparency came to the forefront, prompting researchers to address these concerns within the governance framework.

2024: Emerging topics and future directions

In 2024, emerging themes such as **learning**, **data development**, **global governance**, and **decision-making** are prevalent. The reappearance of **health** and **AI** reflects ongoing research interest in applying data governance to health informatics and AI ethics. Additionally, **global** governance has become a key area of exploration, likely due to the cross-border nature of data flows and the need for international standards. The focus on **review** and **decision-making** suggests a shift toward evaluating existing frameworks and establishing more nuanced decision-making processes in data governance.

This thematic progression demonstrates the adaptive nature of data governance research as it responds to technological advances and societal demands. Early foundational themes provided a base, while recent years highlight the growing intersection of data governance with AI, privacy, and global cooperation. This evolution indicates that future research will likely continue to address ethical considerations, regulatory challenges, and the technical requirements of emerging technologies, ensuring data governance frameworks remain relevant in an increasingly digital and interconnected world.

The advancement of emerging technologies such as blockchain, AI, edge, and quantum computing presents opportunities and challenges for the next phase of data governance research. By offering evolved approaches to every step of the data cycle and addressing existing challenges like data ethics, privacy, security, and transparency, new complexities and ethical considerations also emerge and demand further investigation. Future studies could explore the data governance models integrating ethical AI principles, how AI-driven algorithms impact data governance decision-making processes, examine the integration of the regulatory framework and decentralized systems, and investigate the implications of advanced devices on data sovereignty and cross-border data flows, and so forth. Research gaps not only emerge for each specific technology and its applications, but also reflect evolving cross-cutting challenges. It is important that future research could prioritize adaptive ethical and interoperable frameworks that balance innovation with data protection and individual rights.

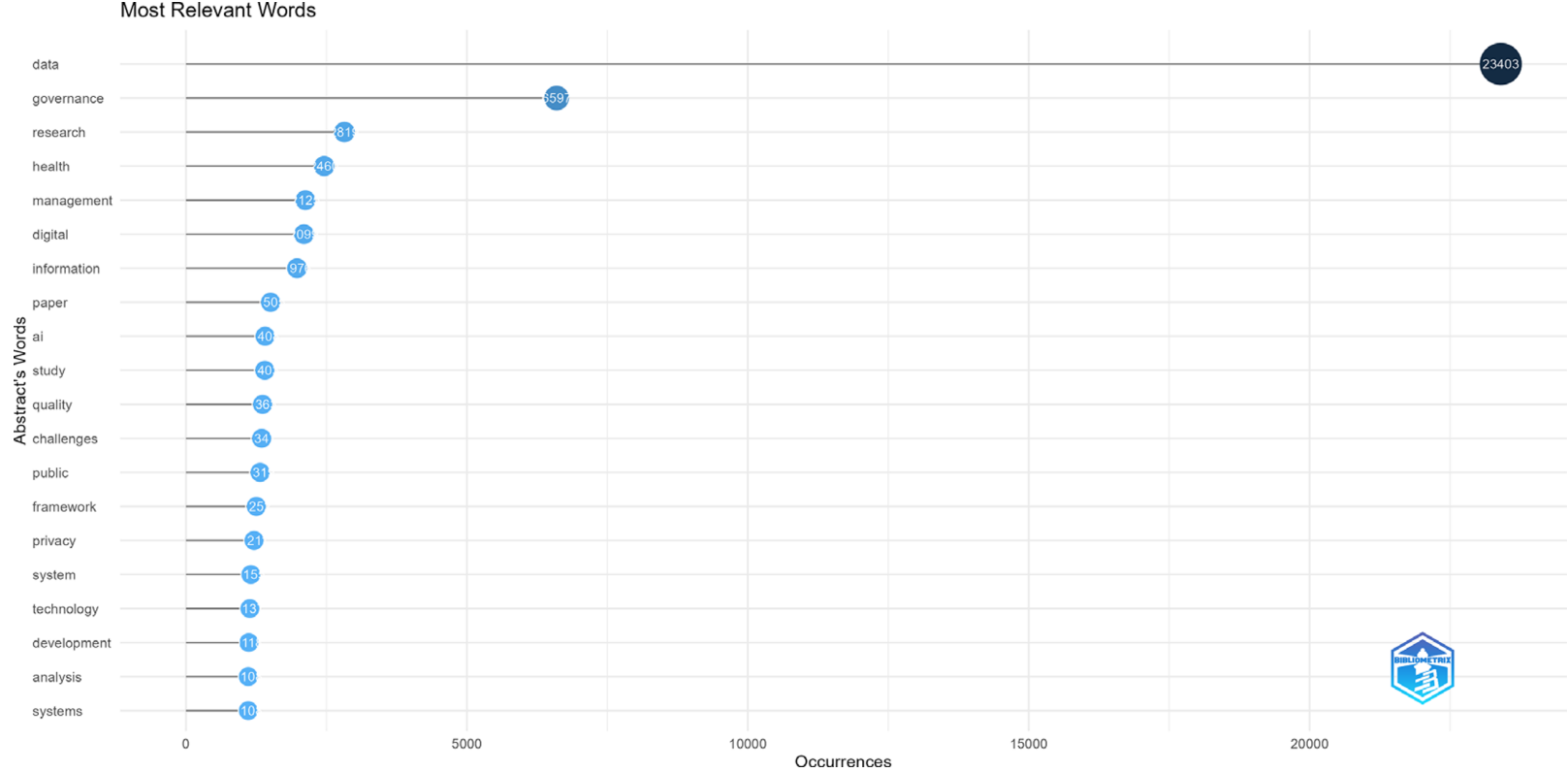


Figure 5. Most relevant keywords in data governance research abstracts.



4. Most relevant in various areas

4.1. Most relevant words in research abstracts

Figure 5 displays the frequency of the most relevant words found in the abstracts of scientific publications, emphasizing the dominant themes and focus areas in recent research.

Key observations:

1. **Prevalence of “Data”:** The word “data” is by far the most frequently mentioned term, appearing over 23,000 times, which indicates a strong emphasis on data-centric research. This trend aligns with the increasing importance of data science, big data, and data-driven decision-making in modern research.
2. **Focus on Governance and Health:** Other frequently occurring terms include “governance” and “health,” with 659 and 481 occurrences, respectively. These words highlight a major focus on research around health systems, governance structures, and policymaking, which may reflect current global challenges and the need for resilient health and governance frameworks.
3. **Emerging Technologies:** Words like “AI,” “digital,” and “technology” suggest an increasing focus on digital transformation, AI, and technological advancements, areas that are becoming integral across multiple research disciplines.
4. **Key Research Themes:** Terms such as “management,” “information,” “quality,” and “challenges” point to research interests in improving management practices, information dissemination, quality control, and addressing various challenges within different sectors.
5. **Privacy and Public Frameworks:** The presence of “privacy,” “public,” and “framework” indicates attention to frameworks that ensure privacy, particularly in data and technology research, and the role of public governance in research application.

This word frequency analysis provides insight into the core themes of recent scientific literature, with a significant focus on data, health, governance, and emerging digital technologies. It reflects the evolving landscape of research, with a notable shift toward data-driven methods and technologies, as well as the exploration of challenges and frameworks within the health and governance sectors.

4.2. Most relevant words in research titles

Figure 6 shows the frequency of the most relevant words found in the titles of scientific publications. The analysis of title keywords provides insight into the primary themes and focal points of recent research.

Key observations:

1. **Emphasis on “Data”:** The term “data” appears most frequently, with 2907 occurrences. This prominence reflects the centrality of data-related research in current scientific discourse, aligning with trends in data science, big data, and data governance as integral aspects of modern research.
2. **Focus on Governance and Health:** Following “data,” terms like “governance” (1322 occurrences) and “health” (341 occurrences) are prevalent. These words suggest significant research interest in governance structures, health systems, and policies, which may correspond to recent global health challenges and an increased focus on public health and regulatory frameworks.
3. **Technological Themes:** Words like “digital,” “AI,” “intelligence,” and “analytics” highlight the emphasis on technology and AI within research. These terms reflect ongoing efforts to explore digital transformations, machine learning, and analytical techniques in various fields.
4. **Management and Frameworks:** The words “management” (267 occurrences) and “framework” (172 occurrences) indicate attention to structuring and managing systems, likely relevant to areas such as organizational studies, governance, and methodological frameworks within research.
5. **Analytical and Review-Based Research:** Terms such as “analysis,” “study,” “review,” and “model” suggest a trend toward analytical approaches and review-based research, indicating a focus on evaluating, synthesizing, and modeling existing knowledge.

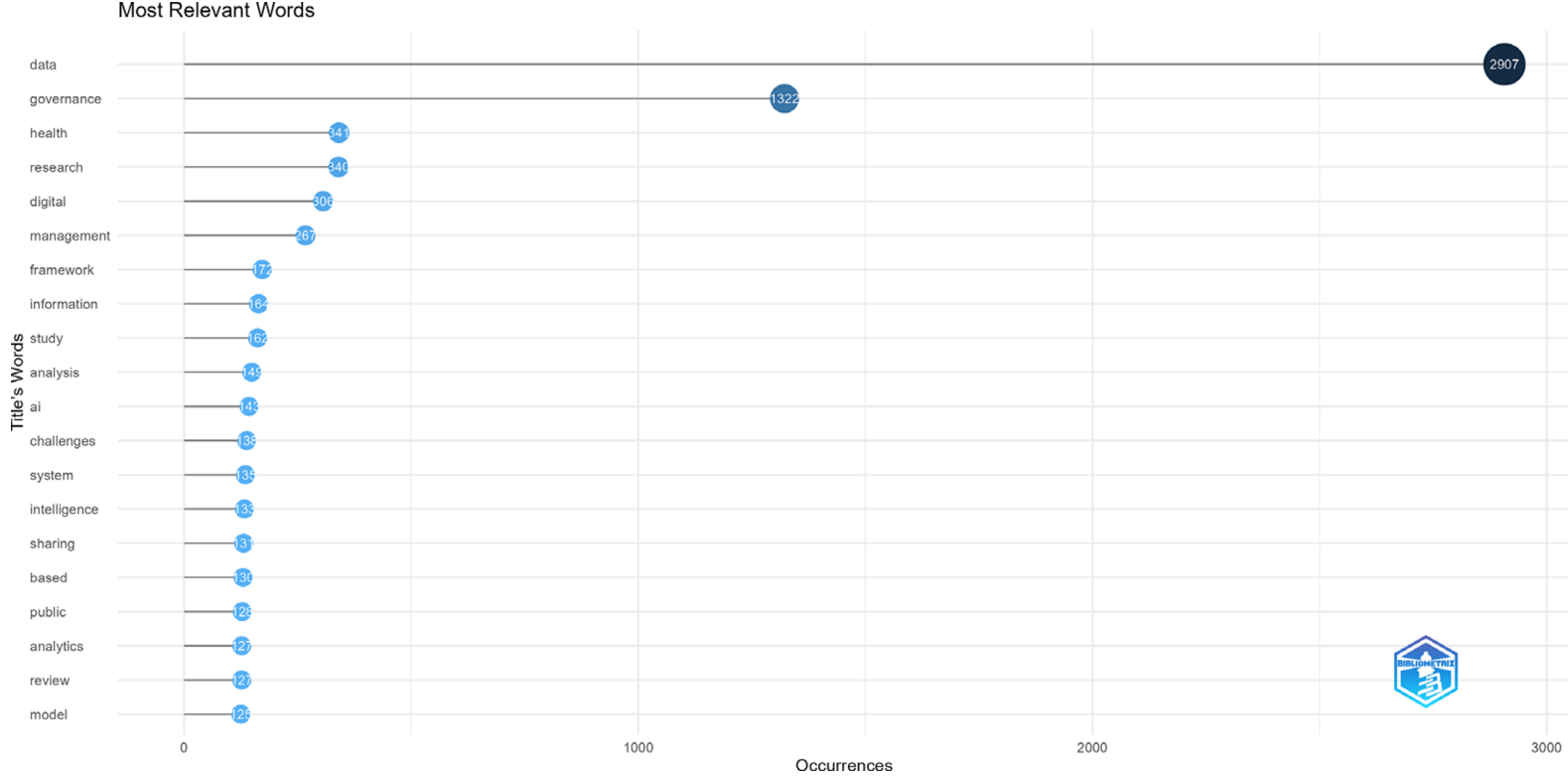


Figure 6. Most relevant keywords in data governance research title.

The prevalence of these keywords in research titles highlights the key areas of interest within the scientific community, with a strong emphasis on data, governance, health, and technological advancements. This distribution points to a research landscape where data-driven methods, health and governance frameworks, and digital innovation are at the forefront of academic and applied research.

4.3. Most relevant affiliations in scientific publications

Figure 7 highlights the most frequently appearing institutional affiliations in recent scientific publications, showcasing the contributions of leading universities and organizations worldwide.

Key observations:

1. **Top Contributor:** The University of Oxford is the most prominent institution, with 40 articles attributed to its researchers. This places Oxford as a major contributor to the global research output, likely reflecting its strong research infrastructure and interdisciplinary focus.
2. **Leading Institutions:** Other major contributors include the University of Toronto, the University of Edinburgh, the Delft University of Technology, and the University of Melbourne, each with over 18 publications. These universities are known for their active research communities and global collaborations, positioning them as influential players in scientific research.
3. **Diverse Geographical Representation:** The affiliations span across multiple continents:
 - **Europe:** Institutions like Oxford, Edinburgh, Delft, and KU Leuven demonstrate strong European involvement in research.
 - **Asia:** The State Grid Corporation of China appears as a significant nonacademic research organization from Asia, underlining its role in research, especially in fields related to technology and infrastructure.
 - **North America:** Stanford University, Harvard University, and the University of Toronto showcase North America's strong presence in research.
 - **Australia:** The University of Melbourne represents Australia's contribution to global research output.
 - **Indonesia:** The University of Indonesia is included, indicating active research participation from Southeast Asia.
4. **Nonacademic Contributors:** The presence of the State Grid Corporation of China reflects the contributions from nonacademic research entities, particularly in fields like engineering, energy, and technology.

Figure 7 illustrates the international nature of scientific research, with contributions from institutions across North America, Europe, Asia, and Australia. It highlights both academic and nonacademic research organizations' roles in advancing knowledge across a variety of fields. The global distribution of these leading institutions underscores the collaborative and interconnected nature of modern scientific research.

5. Publications by various categories and UN SDGs

5.1. Publications by research category

Figure 8 presents the distribution of scientific publications across various research categories according to the Australian and New Zealand Standard Research Classification (ANZSRC) released in 2020, focusing on works that include “data governance” in the title and abstract. This breakdown offers insights into the interdisciplinary nature of data governance research and highlights the fields most engaged with this theme.

Key observations:

1. **Dominance of Information and Computing Sciences:** With 1860 publications, information and computing sciences far outpace other fields. This prominence indicates that data governance is

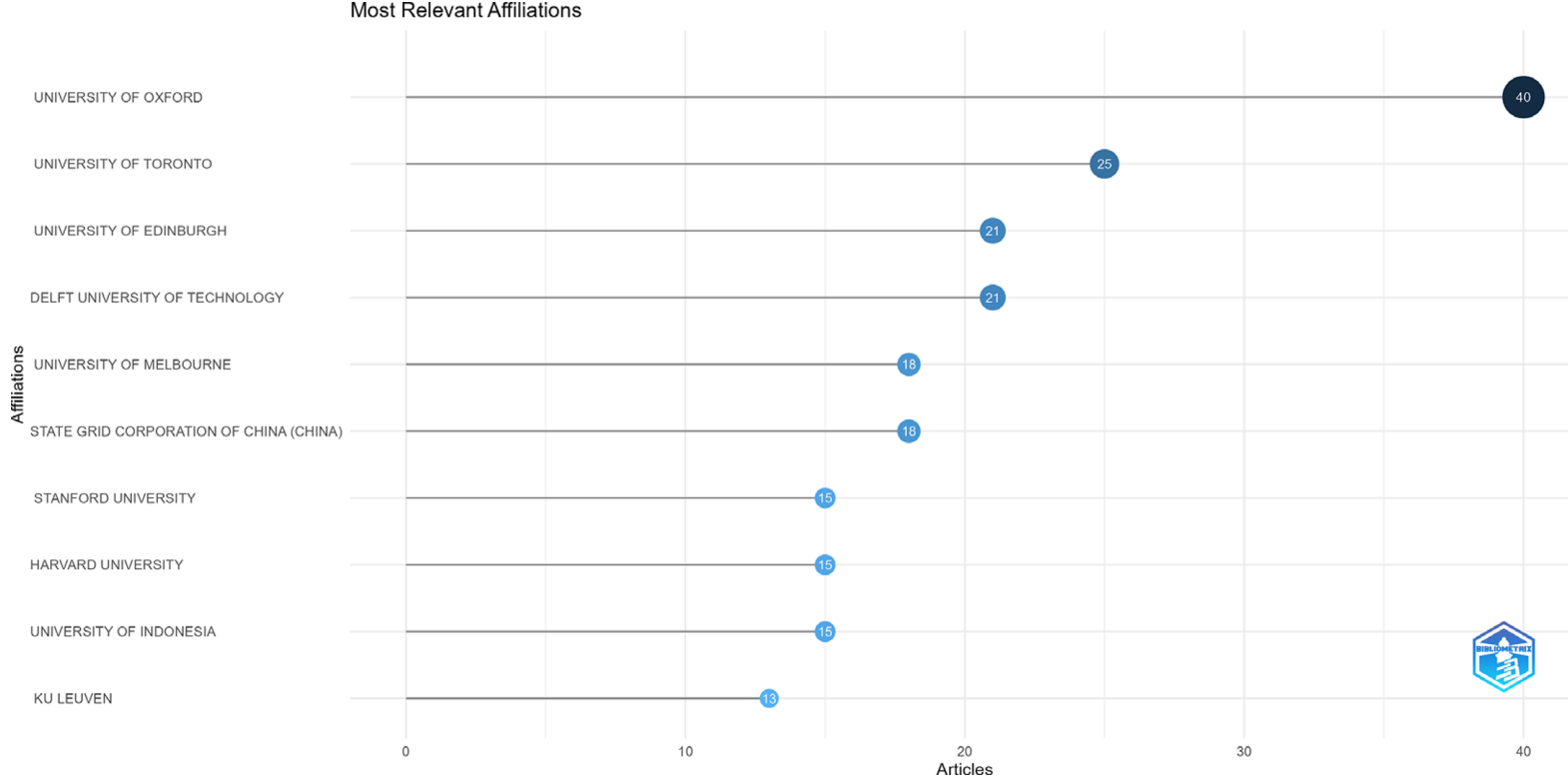
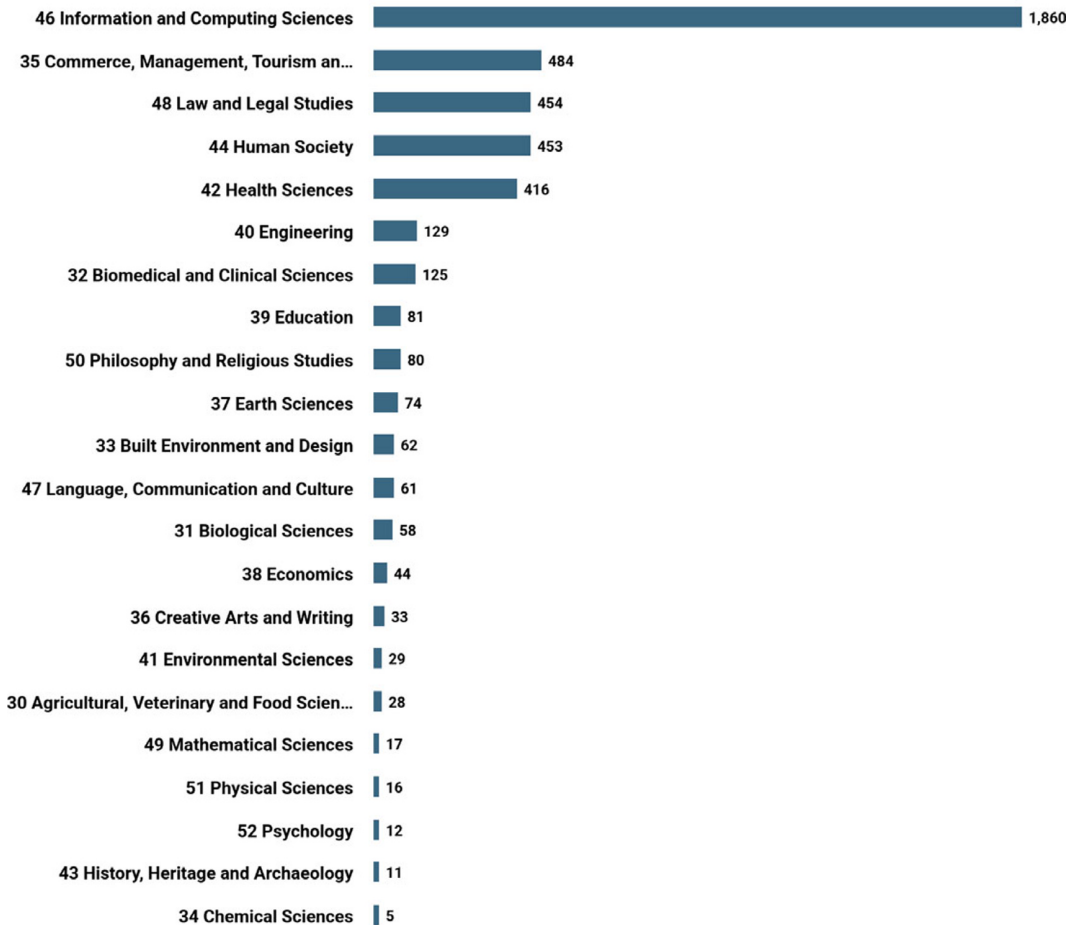


Figure 7. Top institutional affiliations in data governance research publications.



number of publications in each research category. (Criteria: see below)



Source: <https://app.dimensions.ai>

Exported: July 14, 2024

Criteria: "data governance" in title and abstract.

© 2024 Digital Science and Research Solutions Inc. All rights reserved. Non-commercial redistribution / external re-use of this work is permitted subject to appropriate acknowledgement. This work is sourced from Dimensions® at www.dimensions.ai.

Figure 8. Distribution of data governance publications across research categories (ANZSRC 2020).

heavily explored within the context of computing, data science, and information technology, likely due to its relevance in data management, cybersecurity, and ethical data usage.

- Business and Legal Focus:** Commerce, Management, Tourism, and Services (484 publications), along with Law and Legal Studies (454 publications), also show substantial interest in data governance. This suggests that data governance issues are increasingly important in the business sector for compliance, risk management, and regulatory practices, as well as within the legal realm for policy formulation and regulatory frameworks.
- Social and Health Sciences Engagement:** Categories such as human society (453 publications) and health sciences (416 publications) reflect the importance of data governance in societal and health-related contexts. These fields likely address data privacy, ethical concerns, and the protection of sensitive information, especially within public health and social research.
- Moderate Contributions from Engineering and Biomedical Sciences:** Engineering (129 publications) and biomedical and clinical sciences (125 publications) also engage with data

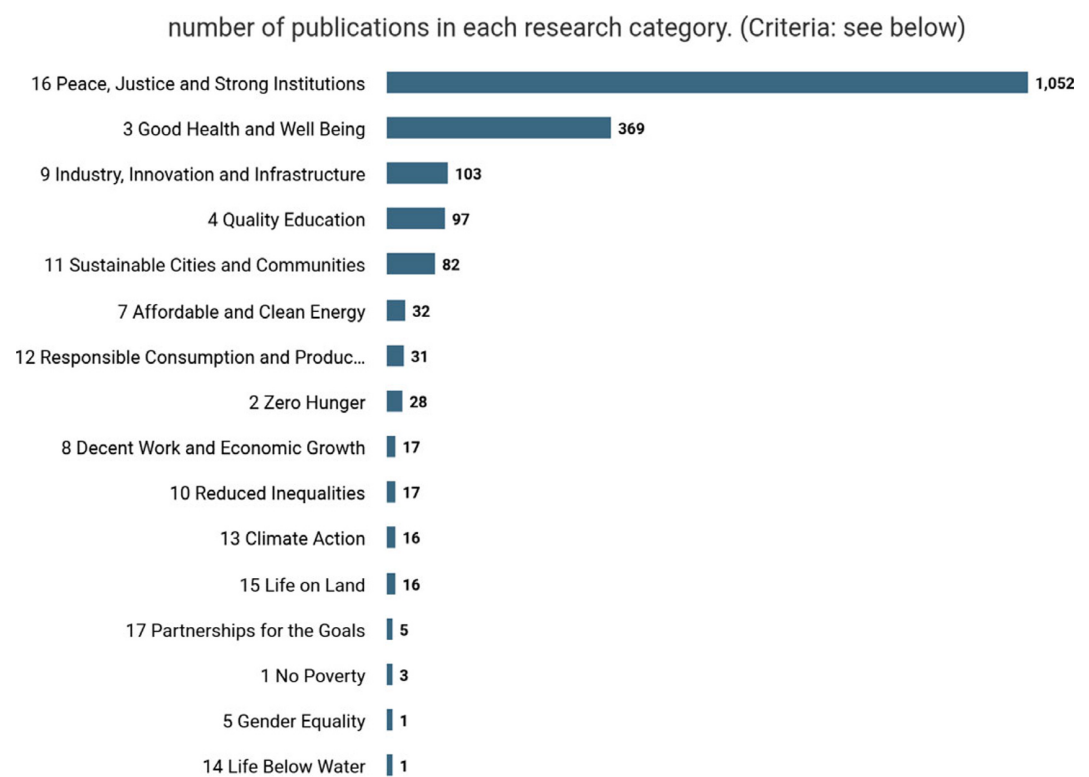
governance, likely focusing on the technical aspects of secure data infrastructure and patient data confidentiality.

5. **Less Representation in Traditional Sciences:** Categories like biological sciences (58 publications), earth sciences (74 publications), and chemical sciences (5 publications) have lower counts. These fields may engage with data governance in more specific contexts, such as data management in ecological studies or chemical data safety, but data governance is not a primary focus.

This distribution underscores the interdisciplinary impact of data governance, with significant contributions from computing, business, law, and health sciences. The varying levels of engagement across fields highlight the importance of tailored data governance strategies, particularly in areas where data privacy, regulatory compliance, and ethical concerns are paramount.

5.2. Publications by research category (aligned with UN SDGs)

Figure 9 shows the distribution of scientific publications across research categories that align with the United Nations Sustainable Development Goals (UN SDGs). This categorization provides insights into how “data governance” is addressed in research focused on sustainable development objectives. It is worth noting that not all publications have clarified how their research is linked to any specific UN SDGs, and there are also publications which address multiple UN SDGs. The sum of occurrences for all UN



Source: <https://app.dimensions.ai>
Exported: July 14, 2024
Criteria: "data governance" in title and abstract.

© 2024 Digital Science and Research Solutions Inc. All rights reserved. Non-commercial redistribution / external re-use of this work is permitted subject to appropriate acknowledgement. This work is sourced from Dimensions® at www.dimensions.ai.

Figure 9. Data governance publications aligned with un sustainable development goals. Data source: Dimensions.

SDGs is 1870 based on the record of 3543 documents. Although there is no available data on how many unique documents have aligned with one or more UN SDGs, this outcome still reflects the importance of data governance research in overcoming global challenges, with uneven distribution of focus across all UN SDGs.

Key observations:

1. **Emphasis on Peace, Justice, and Strong Institutions (SDG 16):** With 1052 publications, SDG 16, which focuses on peace, justice, and building strong institutions, is the most frequently addressed category in data governance research. This emphasis reflects the critical role of data governance in establishing transparency, accountability, and trust within institutions.
2. **Focus on Good Health and Well-Being (SDG 3):** SDG 3, addressing health and well-being, has 369 publications. This significant number indicates an active interest in ensuring data governance in health-related fields, especially in protecting patient data, improving health systems, and managing public health information.
3. **Industry, Innovation, and Infrastructure (SDG 9):** With 103 publications, SDG 9 shows the application of data governance in supporting innovation and building resilient infrastructure. This connection highlights the importance of data management in fostering industrial development and technological advancements.
4. **Attention to Quality Education (SDG 4):** SDG 4, with 97 publications, suggests a role for data governance in educational settings, likely involving data privacy for students, the use of digital platforms, and data-driven educational policies.
5. **Moderate Representation in Other SDGs:** Categories such as sustainable cities and communities (SDG 11) with 82 publications and affordable and clean energy (SDG 7) with 32 publications reflect data governance's relevance in urban planning, sustainability, and energy sectors.
6. **Limited Focus on Goals like Gender Equality and Life Below Water:** Goals such as gender equality (SDG 5) and life below water (SDG 14) have minimal publications (1 each), indicating less focus on data governance in these areas. This may suggest opportunities for further research on data governance in fields like environmental conservation and social equity.

The prominence of SDGs like peace, justice, and strong institutions and good health and well-being underscores data governance's role in promoting institutional integrity and protecting health data. This distribution highlights the alignment of data governance practices with the UN's sustainable development agenda, especially in fields where ethical, transparent, and secure data handling is crucial.

6. Trend analysis

6.1. Word frequency over time in research abstracts

Figure 10 tracks the cumulative frequency of specific terms in scientific publications' abstracts from 2007 to 2024. This visualization illustrates trends in the usage of keywords over time, reflecting shifting research interests and emerging themes in scientific discourse.

Key observations:

1. **Dominance of "Data":** The term "data" shows a sharp and continuous increase, particularly accelerating after 2017, and reaching over 20,000 occurrences by 2024. This exponential rise reflects the growing focus on data-driven research, the proliferation of big data, and advancements in data science across various disciplines.
2. **Growth of "Governance" and "Digital":** Terms like "governance" and "digital" also show noticeable upward trends. The increase in "governance" aligns with the rising importance of data governance frameworks, privacy regulations, and ethical concerns in managing data. Similarly,

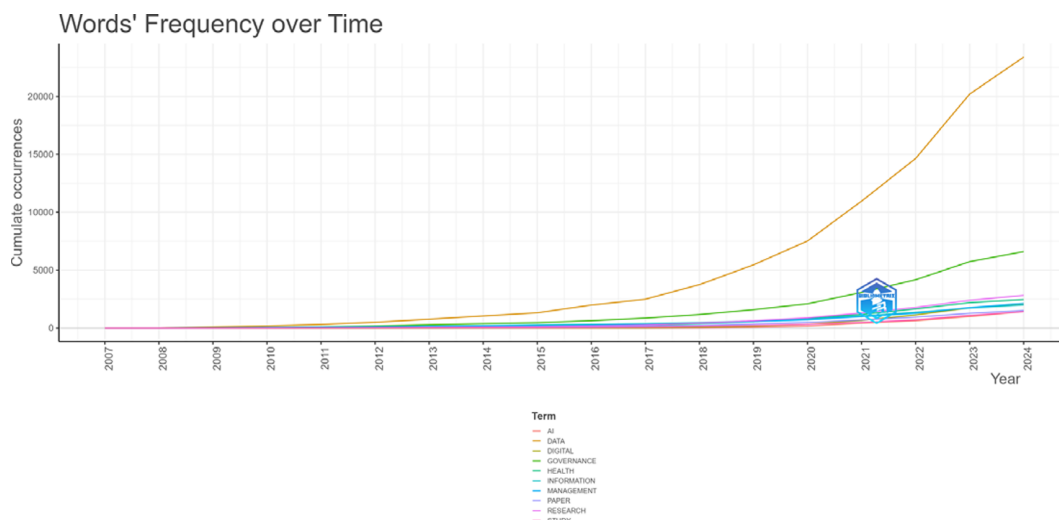


Figure 10. Trends in keyword frequency over time in data governance research (2007–2024).

“digital” reflects the impact of digital transformation and the expansion of digital technologies in research.

- Health-and-Technology-Related Terms:** The terms “health,” “AI,” and “information” also demonstrate steady growth, underscoring the integration of health data, AI, and information management in recent research.
- Emerging Presence of “AI” and “Study”:** “AI” and “study” have lower frequencies than terms like “data” or “governance,” but both have shown steady growth, especially after 2018. This trend highlights the increased application of AI methodologies and a stronger emphasis on research studies focusing on data analytics.
- Other Terms:** Words like “management,” “paper,” and “research” have relatively stable but gradual growth, reflecting consistent interest in general research practices, data management, and scholarly publishing.

The chart reveals that “data” and “digital” terms are the most influential and rapidly growing keywords, indicating a clear shift toward data-centric and digital-oriented research in recent years. The rising usage of “governance” and “AI” further underscores the focus on responsible data use and technological advancements. This trend reflects the increasing complexity and interdisciplinary nature of research, where data governance, health, AI, and digital transformation converge to shape the future of scientific exploration.

6.2. Word frequency over time in research titles

Figure 11 depicts the cumulative occurrences of selected terms in research titles from 2007 to 2024. By tracking these keywords, the chart provides insight into the evolving focus areas and popular themes in scientific research.

Key observations:

- Rapid Growth of “Data”:** The term “data” shows a pronounced increase over time, particularly after 2018, and continues to rise sharply, reaching nearly 3000 cumulative occurrences by 2024. This trend underscores the central role of data-centric research, with “data” emerging as a focal theme across numerous disciplines.
- Significance of “Governance” and “Framework”:** “Governance” and “framework” display noticeable upward trends, reflecting a growing interest in data governance frameworks and

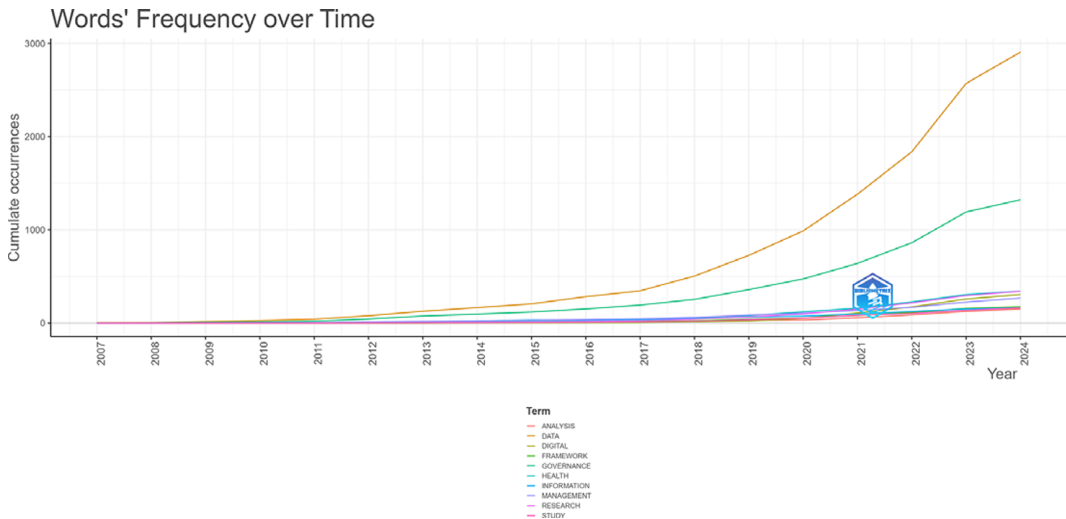


Figure 11. Keyword frequency trends in data governance research titles (2007–2024).

structured approaches for managing data and digital resources. This increase aligns with rising awareness around data privacy, regulatory compliance, and the need for responsible data handling.

3. **Emerging Importance of “Digital”:** The term “digital” has also shown considerable growth, indicating the widespread influence of digital transformation and technological advancements on research topics, especially in fields adopting digital solutions for data management and analysis.
4. **Steady Increase in Health and Information Terms:** “Health” and “information” exhibit steady growth, pointing to ongoing research in health information management, health data protection, and the role of information in public health and healthcare settings.
5. **Consistent Presence of Analytical Terms:** Terms such as “analysis,” “research,” and “study” show consistent growth, indicating a focus on methodological rigor, analytical frameworks, and systematic research approaches across various fields.

The chart highlights “data” as the most influential keyword, with a rapid upward trajectory over recent years, showing the prominence of data-oriented research. The rising frequency of “governance,” “framework,” and “digital” suggests an increasing emphasis on structured data management and digital innovation. This pattern reflects a research landscape where data-driven insights, ethical governance, and digital applications are key components shaping scientific inquiry.

7. Conclusion

This study offers a comprehensive analysis of the evolution of data governance research from 2007 to 2024, drawing from an extensive dataset of over 3500 academic publications authored by more than 9000 researchers. It highlights the significant growth in data governance research, driven by the increased importance of managing data responsibly across sectors. Key themes have emerged over time, reflecting shifts in focus from foundational concepts of data privacy and regulatory compliance to the integration of AI, global governance, and health data.

The findings underscore the interdisciplinary nature of data governance, with high engagement from fields like information and computing sciences, business, health, and law. The global relevance of data governance is further demonstrated by the substantial international collaboration and a high rate of citations, illustrating the field’s impact on academia, policy, and practice. Additionally, alignment with the UN SDGs highlights the contribution of data governance to broader societal objectives, particularly in fostering peace, justice, strong institutions, and health.

It should be noted that the increasing interest in data governance within organizations like the United Nations, G20, and other international forums signals a promising avenue for further research. For instance, conducting a full text search of the keyword “data governance” in *the United Nations Digital Library* would result in 5 documents in 2007, 19 documents in 2018, and 112 documents in 2024, reflecting a 2140% increase since 2007 and a 489% increase since 2018 in policy-focused discourse. As new normative standards emerge from these bodies, the academic community is likely to respond with a growing body of work that examines, critiques, and expands upon these standards. In recent years, there has been a notable rise in conferences dedicated to data governance, such as dedicated sessions hosted by the United Nations Statistical Commission (UNSC) at its 2023 and 2024 annual meetings and the data stewardship initiative launched by the United Nations Economic Commission for Europe (UNECE) in 2022. In 2023, the United Nations System Chief Executives Board for Coordination endorsed “*International Data Governance – Pathways to Progress*” program, which was initially developed by the UN High-level Committee on Programmes (HLCP) international data governance working group, coled by the United Nations Office of Drugs and Crime (UNODC) and the World Health Organization (WHO). Additionally, in 2024, at the Summit of the Future, proposals within the “*Pact for the Future*” emphasize the urgency and global relevance of data governance frameworks. Same year, the United Nations General Assembly requested the Commission on Science and Technology for Development (CSTD) to establish a dedicated *working group on data governance at all levels*, with the new multi-stakeholder working group now established bringing together 27 UN member states, the Global Data Governance (GDG) initiative was launched recently in 2025 and aims to address critical challenges in areas such as cross-border data flows, digital sovereignty, and ethical AI deployment.

This review serves as a foundational benchmark, providing valuable insights for researchers, policy-makers, and practitioners. To build on this comprehensive review, future research can investigate policy documents and their connections with research publications and make data-driven measurements on the actual impact of research on data governance policies. As data governance continues to evolve, future research will likely need to address emerging ethical, regulatory, and technological challenges, ensuring that governance frameworks remain relevant and effective in an increasingly interconnected and data-driven world. The findings of this study also underscore the urgency for policymakers to adopt forward-looking technology-responsive strategies to address governance gaps in an era of rapid technological advancement. Specific recommendations include adopting independent audits of AI systems to mitigate biases and enhance transparency; improving real-time monitoring of data ecosystems through advanced tools to strengthen compliance and security; and fostering multi-stakeholder collaboration to ensure that the governance framework remains inclusive and adaptable. It is also critical to develop interoperable protocols for cross-border data flows, clearly define data ownership and sovereignty, and establish trust and consistency for emerging technologies via certification programs. Moreover, to ensure that frameworks remain adaptive while balancing innovation and data governance challenges, regulatory innovations such as regional or country-level sandboxes for safe and controlled experimentation should be promoted. These can be complemented by efforts to build on established international guidelines to harmonize conflicting national policies, as well as by encouraging and rewarding innovation through mechanisms like case studies, awards, grants, and tax incentives—all of which can foster responsible AI development and cross-border alignment.

Data availability statement. The authors confirm that the data supporting the findings of this study are available within the article.

Author contribution. Investigation: H.H.; X.H.; and S.M.F.; writing, review, and editing: H.H.; X.H.; and S.M.F.

Funding statement. This work received no specific grant from any funding agency, commercial, or not-for-profit sector.

Competing interests. None declared.

References

- Aria M and Cuccurullo C** (2017) bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11(4), 959–975.
- Aria M, Cuccurullo C, D’Aniello L, Misuraca M and Spano M** (2022) Thematic analysis as a new Culturomic tool: The social media coverage on COVID-19 pandemic in Italy. *Sustainability* 14, 3643. <https://doi.org/10.3390/su14063643>.
- European Union** (2022) Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act). Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868> (accessed 29 October 2024).
- Eurostat** (2020) European Statistics Code of Practice. Available at <https://ec.europa.eu/eurostat/documents/3859598/11215121/KS-GQ-20-001-EN-N.pdf/7d38e7b8-48f3-314f-177d-4c5c7c92e6e3> (accessed 30 May 2023).
- G20** (2019) G20 Osaka Leaders Declaration. Available at https://www.mofa.go.jp/policy/economy/g20summit/osaka19/en/documents/final_g20_osaka_leaders_declaration.html (accessed 15 May 2023).
- G7. Carbis Bay G7 Summit Communiqué** (2021) Our Shared Agenda for Global Action to Build Back Better. Available at <https://www.consilium.europa.eu/media/50361/carbis-bay-g7-summit-communique.pdf> (accessed 29 October 2024).
- G7. G7 Roadmap for Cooperation on Data Free Flow with Trust** (2021) G7 Digital and Technology Track—Annex 2. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986160/Annex_2_Roadmap_for_cooperation_on_Data_Free_Flow_with_Trust.pdf (accessed 29 October 2024).
- G77** (2023) Statement on Behalf of the Group of 77 and China by the Delegation of the Republic of Cuba at the Informal Consultations on the Global Digital Compact on the Third Thematic Deep Dive: ‘Data Protection’. New York, 24 April. Available at <http://www.g77.org/statement/getstatement.php?id=230424b> (accessed 30 May 2023).
- Hassani H and MacFeely S** (2023) Driving excellence in official statistics: Unleashing the potential of comprehensive digital data governance. *Big Data and Cognitive Computing* 7, 134. <https://doi.org/10.3390/bdcc7030134>.
- International Monetary Fund (IMF)** (2014) External Sector Report: Statistical Appendix. Available at <https://www.imf.org/external/pubs/ft/scr/2014/cr1404.pdf> (accessed 30 May 2023).
- MacFeely S** (2020) In search of the data revolution: Has the official statistics paradigm shifted? *Statistical Journal of the IAOS* 36, 1075–1094.
- MacFeely S, Me A, Fu H and Schweinfest S** (2020) We urgently need a global data convention. *World Economic Forum—Pioneers of Change Blog* 13 November. Available at <https://www.weforum.org/agenda/2020/11/global-data-convention> (accessed 29 October 2024).
- MacFeely S, Me A, Fu H, Veerappan M, Hereward M, Passarelli D and Schüür F** (2022) Towards an international data governance framework. *Statistical Journal of the IAOS* 39, 703–710.
- Me A, Fu H and MacFeely S** (2021) A global data convention to safeguard sustainable development. *World Data Forum Data Blog* 5 October. Available at <https://unstats.un.org/unsd/undataforum/blog/a-global-data-convention-to-safeguard-sustainable-development/> (accessed 29 October 2024).
- SJR**. Scimago Journal & Country Rank: Country Rankings for all Subject Areas from 1996–2024. Available at <https://www.scimagojr.com/countryrank.php> (accessed 21 April 2025).
- UNCTAD** (2021) Digital Economy Report 2021—Cross-border Data Flows and Development: For Whom the Data Flow. Available at https://unctad.org/system/files/official-document/der2021_en.pdf (accessed 29 October 2024).
- United Nations** (2017) Principles and Recommendations for Population and Housing Censuses: Revision 3. Available at https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M6rev3-E.pdf (accessed 30 May 2023).
- United Nations** (2021) Our Common Agenda—Report of the Secretary General. Available at https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf (accessed 29 October 2024).
- United Nations** (2023) International Data Governance: Pathways to Progress. United Nations System Chief Executives Board for Coordination—Advance Unedited Version of 4 May. Available at <https://unsceb.org/international-data-governance-pathways-progress> (accessed 29 October 2024).
- United Nations Statistical Commission** (2013) Fundamental Principles of Official Statistics. Available at <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf> (accessed 30 May 2023).
- World Bank** (2021) World Development Report 2021: Data for Better Lives. Available at <https://www.worldbank.org/en/publication/wdr2021> (accessed 29 October 2024).
- World Bank** (2020) Data for Better Lives: A Decade of World Development Indicators. Available at <https://openknowledge.worldbank.org/bitstream/handle/10986/34158/9781464815678.pdf> (accessed 30 May 2023).