Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Data governance research and the data economy: Quantitative evidence on thematic imbalance, knowledge diffusion and collaboration

Revyakin Sergey

srevyakin@hse.ru

National Research University Higher School of Economics

Research Article

Keywords: Data governance, data economy, bibliometric indicators, knowledge diffusion, research collaboration, thematic imbalance

Posted Date: October 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10767056/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Data governance has become a prominent interdisciplinary research area, yet the balance between its technological, economic, and organizational dimensions remains insufficiently understood. This study provides quantitative evidence on the thematic structure, knowledge diffusion, and collaboration patterns of data governance research. The analysis covers 2,128 English-language articles, conference papers, and reviews indexed in Scopus for 2012–2024. Descriptive bibliometric indicators, subject-area profiles, SciVal topic metrics, citation links to scholarly publications, policy documents and patents, and country co-authorship networks were examined using Scopus, SciVal, and VOSviewer. The results show sustained growth in publication output, accompanied by a marked concentration in computer science and technology-oriented research. In contrast, research on economic value creation, organizational data use, and information-management practices remains comparatively limited. Citation evidence further indicates that the field is predominantly diffused within academic literature: citations in policy documents and patents are substantially less frequent than scholarly citations. International collaboration is associated with stronger citation performance, but research production and influence remain concentrated in a small group of countries. By integrating thematic composition, diffusion indicators, and collaboration structures, the study demonstrates a persistent mismatch between the composition of the knowledge base and the economic and organizational issues associated with data governance. The findings provide an empirical basis for research on data value, organizational governance mechanisms, and the policy translation of data-governance knowledge.

1. Introduction

Driven by digitalization, organizations and economies generate rapidly increasing volumes of data to support decision-making, service delivery, and innovation (Abraham et al., 2023). Data has emerged as an asset with dedicated commercial models and markets (Driessen et al., 2022; Koetzle et al., 2018), creating new opportunities but also complex challenges for managing, protecting, sharing and using information across organizational and national boundaries.

These developments raise fundamental information management questions. For instance, who owns the data generated by connected vehicles and what information should be retained (Soley, 2018)? Where should data centers be located for multinational businesses and how should cross-border data flows be governed? How can organizations and regulators address misinformation and harmful content (Treverton & Esfandiari, 2021)? Should platform-based delivery companies share operational data with restaurant owners (Goel, 2019)?

Existing studies suggest that data governance policies are predominantly driven by precautionary motives—privacy, security and risk mitigation—while economic growth and well-being remain secondary considerations (Vial, 2023; World Economic Forum, 2020). The tangible consequences of data breaches and misuse are easier to demonstrate than the economic benefits of data reuse, which may partly explain the dominance of risk-focused agendas. This study examines whether the composition and diffusion of the data governance knowledge base are aligned with the economic and organizational questions that data-intensive organizations and public institutions face. The central expectation is that the rapid expansion of technology-oriented research has not been matched by comparable development of research on data value, organizational data use, and governance mechanisms. This expectation is assessed quantitatively through the field's subject-area composition, thematic indicators, citation links to academic, policy and patent documents, and international collaboration patterns.

To examine this issue, the study analyzes data governance-related publications indexed in Scopus between 2012 and 2024 using a quantitative bibliometric design. The analysis integrates four complementary dimensions: publication growth, thematic and subject-area composition, knowledge diffusion across scholarly, policy and patent documents, and international collaboration structures. Together, these indicators make it possible to assess how data governance knowledge is produced, distributed across disciplinary domains, and translated beyond the academic literature.

This study makes three contributions. First, it provides an integrated quantitative profile of data governance research by combining thematic composition, citation-based diffusion indicators, and collaboration-network evidence rather than treating these dimensions separately. Second, it identifies and documents a thematic imbalance between technology-oriented research and research addressing economic value creation, organizational data use, and information-management practices. Third, it uses policy and patent citation indicators to assess the extent to which knowledge produced in the field diffuses beyond academic publications, thereby identifying conditions and gaps relevant to the societal and practical translation of research.

The remainder of the article is organized as follows. The second section contains a literature review on data governance literature. The third section describes the research methodology, while the fourth section provides key results of our statistical analysis of the data governance domain. The fifth section contains the overall discussion and suggests areas for future study. The sixth section provides conclusions and limitations.

2. Analytical framework: thematic composition and knowledge diffusion

2.1 Conceptualizations of data governance

There is no consensus on the definition of data governance (Benfeldt et al., 2020). In the context of information management, data governance is commonly understood as the system of policies, processes and responsibilities that determine how data are collected, managed, accessed, shared and used within and across organizations (Table 1).

Table 1
The notions of data governance

Data governance is	Description	Author
Standard design and standard implementation	The organization and implementation of policies, procedures, structure, roles, and responsibilities, which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.	(Ladley, 2010)
Standard (a system of decision rights and accountabilities, framework, set of rules)	A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. The framework that forms the basis for dealing with and managing data for all internal and external stakeholders of an organization. A set of procedures and guidelines that detail how data is to be properly managed, accessed, and used.	(Vial, 2023) (Engels, 2019) (Atlan, 2023)
Enforcement	The use of authority combined with policy to ensure the proper management of information assets.	(Ladley, 2020a)
(Source: Scopus)		

The absence of a single definition for data governance has not prevented it from becoming a popular research topic (Engels, 2019). Although it is still under-researched (Vial, 2023) and the literature points out that the knowledge of data governance remains largely conceptual, which limits its direct applicability to organizational practice (Ransbotham & Kiron, 2017). Scholars still try to estimate the contribution of data to economic growth and well-being (OECD, 2022), while the best practices of data governance are rare (Engels, 2019). At the same time, such practices are highly demanded by practitioners. Moreover, it remains unclear what the most appropriate approach is to data governance (Treverton & Esfandiari, 2021) because data as the subject of the governance is perceived to be a nebulous thing (Singh & Gurumurthy, 2021a), being big, linked, varied, dynamic and streamed at high speeds in real-time (Janssen et al., 2020). Companies also consider data as a valuable asset, but rarely apply the same procedures as to other assets (Engels, 2019), and in many cases it is not easily measurable or accounted for (Ladley, 2020b). Maybe that is one of the reasons why there are still relatively few documented data governance initiatives with clearly measured business or societal outcomes (Ladley, 2020c), which makes it challenging to demonstrate the economic impacts of data governance.

Meanwhile the potential economic, societal, and environmental benefits of facilitating data exchange in the economy cannot be denied (Abraham et al., 2023). However, whether it would be achieved would depend on the institutional framework, which determines the extent to which they can be effectively created, shared, and used (OECD, 2022).

2.2 Governance motives and the data economy

There are three types of reasons why countries engage in data governance: mitigation of potential risks (associated with the use of personal data), productivity growth and cross-border data flow regulation (OECD, 2022). The last two reasons we would consider jointly as the economic motives (Table 2).

Table 2
Summary of data governance methods

Motive for Data governance	Remedies type	Governance type	Article
Precautionary motive	ex post	Governance of personal data use	(Singh & Gurumurthy, 2021a), (Revyakin, 2022)
	ex post	Governance of data of platform-dependent business	(Singh & Gurumurthy, 2021a), (Harracá & Gawer, n.d.)
	ex post	Governance from a competition point of view (antitrust laws)	(Singh & Gurumurthy, 2021a), (UNCTAD, 2021), , (Bloch-Wehba et al., 2019), (Treverton & Esfandiari, 2021)
	ex post	Governance on geoeconomic concentration of data	(Singh & Gurumurthy, 2021a), (UNCTAD, 2019)
	ex post	Governance on cybersecurity (cyber espionage and cyberwarfare)	(Treverton & Esfandiari, 2021)
Economic growth and well-being motive	ex ante	No governance of data trading to properly utilize their value (data exchange)	(Hao et al., 2023), (Muschalle et al., 2012)
	ex ante	No governance of emerged Data marketplaces (data exchange)	(Singh & Gurumurthy, 2021a) (Koetzle et al., 2018) (Driessen et al., 2022) (Stahl et al., 2016) (Abraham et al., 2023)
	ex ante	No international governance on global digital trade (cross-border data flow)	(Treverton & Esfandiari, 2021)

Literature points out that precautionary motive dominates the agenda of data governance (at the expense of economic growth and well-being). This is because unintentional disclosures of sensitive data, a lack of control over data usage, and issues related to trust arising from data sharing have more tangible consequences than potential benefits from related economic growth (World Economic Forum, 2020). That is why more often the existence of data governance is explained from harm prevention perspective (Singh & Gurumurthy, 2021a), addressing privacy and security issues (Revyakin, 2022), rather than from the perspective of unlocking the economic potential of data to foster economic growth and improve well-being of the citizens.

By contrast, economic approaches to data governance – particularly those focused on productivity growth, cross-border data flows, and value creation from data – remain comparatively underdeveloped in both policy and research. At the same time, related streams of literature do exist, including work on the economic value of data, data market design, and data exchange mechanisms (Castellanos et al., 2012; Driessen et al., 2022; Hao et al., 2023). Moreover, some of the risks associated with data use may be mitigated through governance arrangements that not only reduce uncertainty but also create new economic opportunities (Abraham et al., 2023).

2.3 Analytical expectations

The study proceeds from three analytical expectations. First, because data governance has developed alongside rapid advances in digital infrastructure, analytics, artificial intelligence, and cybersecurity, publication output is expected to be concentrated in technology-oriented subject areas. Second, if research on the economic and organizational use of data is comparatively underdeveloped, terms and subject areas related to business, management, economics, and organizational information use should occupy a smaller position in the knowledge base. Third, if data governance knowledge is primarily consolidated within academia rather than translated into applied settings, scholarly citations should exceed citations in policy documents and patents. International collaboration is examined as a complementary indicator because it may be associated with greater visibility and citation impact.

3. Methodology

3.1 Data sources

Scopus was selected as the primary data source because it is a large multidisciplinary database widely used in bibliometric research to analyze the evolution and structure of scientific fields (Arcila-Diaz et al., 2024; Elsevier, 2023). Its broad coverage of peer-reviewed literature makes it appropriate for mapping research trends in the multidisciplinary domain of data governance.

Data governance encompasses a wide range of concepts, including all types of governance under precautionary motive and governance on economic growth and well-being motive. In order to find relevant keywords we analyzed Data Governance topic section (used “Related terms” chapter) of ScienceDirect Topics portal. To identify relevant search terms, the study combined terms derived from prior literature on data governance with iterative refinement using Scopus keyword suggestions and manual screening of highly relevant publications in the field. In addition, Scopus search engine (Keywords section) was used to explore additional keywords. Hence, we selected all relevant keywords and used them to search in the titles, abstracts, or keywords of Scopus database documents, as indicated in Table 3. The search period from 2012 to present was selected to ensure that the paper sample remained up to date.

Table 3
Search query

Description	Condition	Documents
Search query	TITLE-ABS-KEY ("data governance" OR "governance of data" OR "data governances" OR "big data governance" OR "Data Stewardship") AND PUBYEAR > 2011 AND PUBYEAR < 2024	2,626
Search query after refine	TITLE-ABS-KEY ("data governance" OR "governance of data" OR "data governances" OR "big data governance" OR "Data Stewardship") AND PUBYEAR > 2011 AND PUBYEAR < 2024 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "re"))	2,128
Access	We included both open access and non-open access documents.	
Search date	22 December 2023	
Publication-year filter	2012–2024[2]	
Subject area	All	
Source type	We limited our search to articles, conference papers and reviews	
Language	We limited our search to the English language	
(Source: Scopus)		

This search strategy ensures that the dataset is up to date, focuses on peer-reviewed scholarly work, and can be replicated by other researchers interested in data governance and information management.

The process applied for delimiting literature is demonstrated on the Fig. 1.

The preliminary search retrieved 2,626 documents on 22 December 2023. The search applied a publication-year filter of 2012–2024. The records assigned to 2024 represent online-first and article-in-press items available in Scopus on the retrieval date and should not be interpreted as complete 2024 annual output. Most of them were articles and conference proceedings in English. We selected three types of publications in English for further analysis (articles, conference papers and reviews), because they are the typical categories for bibliometric analysis (Hao et al., 2023). Therefore, 2,128 papers were selected for further analysis.

3.2 Research methodology

A bibliometric review is well suited to a multidisciplinary field such as data governance because it enables the systematic analysis of publication output, thematic development, citation patterns, and collaboration structures across a large body of literature (Hao et al., 2023). Bibliometric methods rely on

quantitative analysis of bibliographic data from scholarly abstract and citation databases to identify research trends, emerging topics, and global research networks (Ellili et al., 2023). In this study, Scopus provides the bibliographic data, while VOSviewer and SciVal are used for mapping and visualizing the scientific landscape of data governance.

The study uses a descriptive quantitative bibliometric design. Publication counts were used to characterize field growth; Scopus subject-area classifications and SciVal topic/keyphrase indicators were used to assess thematic composition; citations in scholarly publications, policy documents, and patents were used as differentiated indicators of knowledge diffusion; and co-authorship links between countries were used to characterize international collaboration. Scholarly citations are interpreted as academic diffusion, citations in policy documents as an indicator of policy uptake, and patent citations as an indicator of technological or commercial uptake (Qiao et al., 2025). These indicators do not establish direct causal effects or the quality of individual research outputs; rather, they provide complementary evidence on the production, circulation, and external uptake of knowledge in the field.

Along with that, we analyzed the structure of publication by countries and their collaboration network to conclude on the most influential countries. We defined the most influential countries in data governance domain as the countries that are leading by generation and use of knowledge in the field. To make such conclusion we analyzed countries' publication activity, citation per article, citations in policy document, the location of patent owners who cite the articles, patent offices location, most collaborative countries in the domain and the countries with the biggest number of publication partners.

3.3 Data treatment and comparability

All indicators were calculated at the document level using the metadata and classifications available in Scopus and SciVal at the time of data retrieval. Because SciVal indicators have different update cycles and availability windows, the reporting period for each indicator is stated in the respective table or figure. Results based on scholarly citations, policy-document citations, patent citations, and field-weighted citation impact are interpreted separately and are not aggregated into a single composite measure. A document may contribute to more than one country because it can include authors with multiple national affiliations..

4. Statistical analysis of data governance literature: key results

4.1. Publication growth and general trends analysis

The growth of data generation and data use worldwide has increased the importance of governance arrangements capable of coordinating access, exchange, and use across organizational and regulatory settings (Abraham et al., 2023; Andreev et al., 2023). One indicator of how science responds to this development is the growth of publications addressing data governance.

The number of publications over time is an important indicator of the development of scientific domain because it is useful to analyze the trends of a field (Ellili et al., 2023; Hao et al., 2023). Using the statistics of publications dataset (bibliographic data, including the number of publications and the frequency of citations), we assessed the general development trend of data governance research.

We found that the total number of data governance papers that were published between 2012 and 2024 is 2,128 documents (Fig. 2). In the field of data governance, **there is an overall polynomial increase in the number of publications**, and R^2 of the fitted curve of citation frequency is 0.9766 that allows predicting further growth.

As Fig. 2 shows, there is a definite trend in scientific articles being published that have a term related to data governance in their title, abstract or keywords. In 2012, only 36 documents were published, and started to increase on a yearly basis and reached a maximum of 431 documents in 2023. The 2024 count is incomplete because the dataset was retrieved before the end of that publication year; it is retained to capture online-first records assigned to 2024 but is not used to infer the annual trend. Thus, the overall trend is upward and it is expected to grow even further because further development of data-related technologies is still expected, as long as an increase for data generated and the demand for it (Treverton & Esfandiari, 2021) – and defined trend line supports this hypothesis.

Exploring the roots of use of data governance knowledge we analyze the citation structure. We can observe that most of the published articles of the domain have been cited in other scientific publications, and much less - in policy documents and patents (Table 4).

Table 4
The citation metrics of the publication dataset on data governance

Number of articles	**Attribute, #**
Total, 2012–2023: 2,128	Citations*: 6,545
Cited in patents: 39	Patents**: 118
Cited in policy documents: 255	Citing policy documents***: 645
* Number of citations. (Source: Scopus)	
* Number of citing patents. (Source: SciVal)	
* Number of citing policy documents. (Source: SciVal)	

As Table 4 shows, data governance knowledge is used much more extensively in scholarly publications than in policy documents or patents. This pattern suggests that the field is currently more strongly oriented toward academic knowledge accumulation than toward direct translation into policy or business

practice. The finding is broadly consistent with prior studies that describe data governance knowledge as conceptually rich but still limited in terms of practical implementation guidance (Engels, 2019; Ransbotham & Kiron, 2017).

4.2. Data governance research trends

The subject-area structure of the dataset (Fig. 3) indicates that data governance research is dominated by technology-oriented and health-related fields, especially Computer Science, while Business, Management and Accounting, as well as Economics, Econometrics and Finance, remain comparatively less represented.

At the same time, when we look at the dimensions of frequently used keywords in the dataset, it seems that **economic terms are growing**. (Table 5). Among the top 10 keyphrases prioritized by relevance Ethics, Digital Transformation and Data Access dominate the ranking of terms by scholarly output growth rate.

Table 5
Top 15 keyphrases scholarly output (by growth % over the period 2018–2023) of data governance research

1. Ethics	**Scholarly Output (growth %. over the period 2018–2022)**	**Total # of publications in 2018–2023**	**Relevance (max value = 1.00)**
	1900.0	**169**	**0.06**
2. Digital Transformation	1700.0	69	0.03
3. Data Access	1300.0	77	0.04
4. Intelligence Data	1300.0	32	0.02
5. Stewardship	1100.0	41	0.02
6. Digital Data	1000.0	46	0.03
7. Information Management	866.7	178	0.07
8. Smart Cities	675.0	65	0.03
9. Sovereignty	666.7	57	0.03
10. Social Responsibility	600.0	79	0.04
(Source: SciVal)			

This pattern suggests that the field may be gradually expanding from risk-oriented governance concerns toward issues related to data use, digital transformation, and value creation. From a quantitative social-research perspective, this shift indicates that the thematic composition of the field is changing, although technology-oriented research continues to dominate its publication base. The pattern identifies a measurable gap between the expansion of the field and the relative representation of research on organizational data use and economic value creation.

4.3 Publications in major countries and regions

Table 6 shows that institutional collaboration is the most common form of co-authorship in data governance research (32.3%), followed by national (26.2%) and international collaboration (25.6%), while single-authored papers account for 15.7%. This pattern suggests that data governance research is still more frequently conducted within institutions and countries than across borders.

Table 6
Geographical collaboration for the data governance research dataset (2018–2024)

Metric		**Scholarly Output**	**Citations**	**Citations per Publication**	**Field-Weighted Citation Impact**
International collaboration	25.6%	453	6784	15	2.23
National collaboration	26.2%	464	3613	7.8	1.15
Institutional collaboration	32.3%	572	2886	5	0.97
Single authorship (no collaboration)	15.7%	278	1375	4.9	0.87
(Source: SciVal)					

As shown in Fig. 4, the United States, the United Kingdom, and China dominate the data governance research landscape in terms of publication output, with other highly digitalized economies such as Germany, the Netherlands, Canada, Australia, Italy, Switzerland, and Spain also contributing substantially.

Note

The article could have two or more affiliation countries

As indicated by the field-weighted citation impact and view metrics in Table 7, most of these countries achieve citation performance above the world average, suggesting that highly productive countries also tend to produce influential and visible work in data governance.

Table 7
FWCI, Citation and views per article for data governance research dataset (2018–2023)

Country/Region	Field-Weighted Citation Impact	Citation per article	Views per article
United States	1.85	13	38
United Kingdom	2.16	17	49
China	0.86	3	24
Germany	1.92	11	40
Netherlands	2.42	17	38
Canada	2.22	12	31
Australia	1.84	15	40
Italy	1.94	14	44
Switzerland	3.1	21	54
Spain	1.63	8	65
(Source: SciVal)			

26% of publications co-authored with institutions in other countries/regions (Source: SciVal). The network of international collaboration in the area of data governance is shown in Fig. 5.

Table 8 shows that highly productive countries tend to collaborate intensively with one another, forming dense international networks in data governance research. The strongest collaboration ties are observed between the United States and the United Kingdom, followed by partnerships such as the United States–Canada and Germany–United Kingdom, which jointly shape the core of the international collaboration network in this field.

Table 8
Collaborative pairs of countries for data governance research dataset

Country #1	Country #2	Number of joint publications
United Kingdom	United States	52
Canada	United States	48
Germany	United Kingdom	36
Italy	United Kingdom	31
Australia	United States	30
Germany	United States	30
Australia	United Kingdom	29
Germany	Netherlands	29
Netherlands	United Kingdom	29
Canada	United Kingdom	25
Belgium	Italy	23
France	United Kingdom	21
Germany	Switzerland	20
Netherlands	United States	20
Switzerland	United Kingdom	20
France	Germany	19
China	United States	18
Germany	Italy	18
Australia	Canada	16
Spain	United Kingdom	16
(Source: VOSviewer on Scopus data)		

At the same time, the United States, the United Kingdom and Italy have the biggest number of partners among the countries as shown in Table 9. Having many collaborative countries, this collaboration is very intensive (the density is rather high). At the same time Japan and Germany, for instance, have a

similar number of collaborative countries (28 and 33, respectively), but significantly differ in intensity of collaboration with them (53 and 284 joint publications, respectively).

Table 9
Collaborative network of highly productive countries

Country	Number of collaborative countries	Density of collaboration	Number of documents
United States	45	391	533
United Kingdom	45	424	311
Italy	38	245	106
Netherlands	38	257	154
Australia	37	210	135
Switzerland	37	178	77
France	35	206	65
Spain	34	177	69
Canada	34	200	147
Belgium	33	175	64
Germany	33	284	173
Japan	28	53	19
Denmark	27	101	36
Greece	26	71	24
Norway	26	100	43
Sweden	26	104	31
Austria	25	80	34
Finland	25	92	43
Portugal	25	57	32
Brazil	23	53	40
Poland	22	45	23
Ireland	22	73	57
South Africa	22	82	37
Iran	21	22	5
China	21	61	268
India	20	49	52
Russian Federation	19	24	16
(Source: VOSviewer on Scopus data)			

As shown in Fig. 5, in terms of the number of documents the United States, the United Kingdom and China dominate the data governance research landscape.

5. Discussion

Data governance has become increasingly important as economies and organizations expand their capacity to generate, combine, and use data (Vial, 2023). Yet policy and research agendas continue to be shaped primarily by precautionary concerns, especially privacy, security, and harm prevention. Prior literature suggests two reasons for this pattern: first, the negative consequences of data misuse are more visible and immediate than the benefits of data-driven growth; second, technological change often outpaces the development of governance frameworks (Singh & Gurumurthy, 2021b; Treverton & Esfandiari, 2021). The present study provides quantitative support for a third explanation: the knowledge base itself is unevenly distributed across disciplinary and thematic domains. The results indicate that technology-oriented research has expanded more rapidly than research addressing economic value, organizational data use, and information-management practices.

The bibliometric analysis supports this interpretation by showing a sustained increase in data governance publications between 2012 and 2024, accompanied by a marked concentration of output in technology-oriented subject areas.

This conclusion is reinforced by the citation patterns summarized in Table 4. Data governance publications are cited far more frequently in academic articles than in policy documents or patents, indicating that knowledge in the field is primarily reused within the scientific community rather than in regulatory or commercial contexts. For information management and policy practice, this suggests that existing research may not yet provide sufficient guidance on economic issues related to data generation, storage, exchange, and use. These findings are consistent with prior work that characterizes data governance as conceptually rich but still under-researched from the perspective of practical implementation and measurable outcomes (Engels, 2019; Ransbotham & Kiron, 2017; Vial, 2023).

Looking at the structure of the dataset by subject area, we found that technology-driven publications (Computer science) dominate the research agenda, as shown in Fig. 4 and Table 5, while there are far fewer publications in the fields like Business, Management, Accounting, Economics, Econometrics and Finance. Therefore, the knowledge in data governance domain mostly address technical issues rather than economic ones. At the same time, we found an increase in publications on the terms as ethics, digital transformation, and data access that clearly demonstrates a shift towards data use topic in research to unlock the economic potential of data (in comparison to harm prevention agenda). Further development of this trend could help better align academic research with the needs of practitioners. It could be a possible direction for academics to intensify their efforts to fill existing knowledge gaps. Other possible directions of economic research of data governance could address the key topics of strategic documents on data governance adopted by some national governments (Fig. 6) (Andreev et al., 2023).

5.1 Implications for quantitative research and practice

The observed thematic imbalance has implications for both quantitative research and practice. For research, it indicates the need for designs that measure how concrete governance arrangements relate to data quality, data sharing, cross-functional coordination, investment in data infrastructure, and value realization. Future studies could combine bibliometric evidence with surveys, organizational case studies, panel data, or comparative policy analysis (Fig. 6) to test these relationships directly. For practice, the results suggest that managers and policymakers have a comparatively limited cumulative evidence base when they seek to evaluate the economic and organizational outcomes of data governance arrangements.

Note

Built on an array of strategic documents in the field of data economics (strategies, programs, plans, policies, etc.) for 2018–2023 by country, including the European Union, UK, Germany, USA, China, India, Singapore, South Korea, Japan, Australia, Canada, Spain, etc., using a national big data analytics system.

Based on (Andreev et al., 2023), we can suggest the following directions of research that could contribute to unlocking the economic potential of data (Table 10).

Table 10
Key policies provided in strategic documents on data governance adopted by some national governments

Motive for Data governance	Topics of research, corresponding with the key areas of regulation
Precautionary motive	1. The dilemma of "data protection and privacy, depersonalization and declassification" vs. "benefit for data consumers"
Economic growth and well-being motive	2. The economic value of data: approaches to assess and standards to implement
	3. Investments in data
	4. Data quality (relevance, timeliness, completeness)
	5. Data culture competencies in the public sector and business
	6. Rules for data exchange throughout the entire lifecycle (producers, integrators, suppliers, users)
	7. Open data management, coordination between public bodies and data stewardship functions
	8. Reuse of data, especially those generated by the private and the public sector
	9. Artificial intelligence as a data consumer
	10. The infrastructure for data collection, storage, and exchange (Internet of Things, communications, digital platforms, data centers, etc.)
Source: (Andreev et al., 2023)	

Table 11 summarizes these patterns across multiple bibliometric indicators. The United States and the United Kingdom consistently rank among the top countries by publication activity, citations in policy documents and patents, citation impact, and collaboration intensity, confirming their leading role in generating and disseminating data governance knowledge.

Table 11
Top three influential countries in data governance based on bibliometric indicators

By publication activity	By citing publications in Policy Documents[4]	Patents owner location	Patent offices where registered[5]	By Citation per article	Most Collaborative countries	With the biggest number of publication partners
US	EU	**US**	**US**	CH	**UK – US**	**US**
UK	**US**	CAN	CN	NL	CAN – **US**	**UK**
CN	**UK**	LUX	**UK**	**UK**	DE –**UK**	IT
Note: CAN – Canada; CH – Switzerland; CN – China; DE – Germany; IT – Italy; LUX – Luxemburg; NL – Netherlands; UK - United Kingdom; US - United States.						
(Source: SciVal)						

6. Conclusion

This study shows that the rapid expansion of data generation and use has created a growing demand for governance arrangements that address not only protection and compliance but also value creation from data. However, the research landscape remains dominated by precautionary and technology-oriented perspectives, while economic and information management-oriented studies remain comparatively limited. This imbalance may partly explain why the economic potential of data is still insufficiently supported by applicable knowledge for organizations and policymakers.

The analysis indicates that publication growth in data governance has been driven primarily by technological research. At the same time, the increasing visibility of terms such as digital transformation, information management, and data access suggests a gradual shift toward questions of data use and value creation. The study also shows that data governance publications are far more frequently cited in academic literature than in policy documents or patents, indicating that the field's practical translation remains limited.

This study offers quantitative evidence that data governance research has expanded rapidly but remains uneven in its thematic composition and external diffusion. Technology-oriented research constitutes the largest share of the field, whereas economic value creation, organizational data use, and information-management practices are comparatively less represented. The differentiated citation evidence also suggests that data governance knowledge circulates primarily within scholarly communication, with more limited visibility in policy documents and patent.

By combining publication trends, thematic indicators, citation-based diffusion measures, and collaboration-network data, the study provides a replicable approach for examining the development and societal translation of interdisciplinary research domains. Future work should extend the analysis with multiple bibliographic databases, transparent longitudinal retrieval windows, and mixed-method designs that connect bibliometric patterns to organizational and policy outcomes.

Declarations

Author Contribution

S.A. Revyakin is the sole author of this manuscript and was responsible for all aspects of the research, including conceptualization, methodology, data collection and analysis, and writing of the manuscript.

Acknowledgement

The author thanks Mikhail Revyakin and Dirk Meissner for their helpful comments on earlier drafts of the manuscript.

Data Availability

The bibliometric dataset analyzed in this study was extracted from Scopus and SciVal (Elsevier) on 22 December 2023 and is subject to the licensing terms of these databases. Due to these licensing restrictions, the raw dataset cannot be made publicly available. Aggregated data, derived indicators, and the search strategy used to construct the dataset are described in the manuscript (see Methodology, Section 3) and can be shared by the corresponding author upon reasonable request, subject to the terms of the Scopus and SciVal license agreements.

References

1. Abraham, R., Schneider, J., vom Brocke, J.: A taxonomy of data governance decision domains in data marketplaces. Electron. Markets. **33**(1) (2023). https://doi.org/10.1007/s12525-023-00631-w
2. Andreev, M., Zinina, T., Sokolova, E.: *Data economy: global approaches to governance*. (2023)., October 12 https://issek.hse.ru/news/865612618.html
3. Arcila-Diaz, J., Delgado-Caramutti, J., Millones-Gómez, P.A., Figueroa-Quiñones, J., Valencia-Arias, A.: Research trends in Peruvian universities: proposal for a research agenda with a bibliometric approach. Scientometrics. (2024). https://doi.org/10.1007/s11192-024-05211-z

4. Atlan: *Data Governance vs Information Governance: What's Your Priority?* Atlan. (2023)., May 18 https://atlan.com/data-governance-vs-information-governance/
5. Benfeldt, O., Persson, J.S., Madsen, S.: Data Governance as a Collective Action Problem. Inform. Syst. Front. **22**(2), 299–313 (2020). https://doi.org/10.1007/s10796-019-09923-z
6. Bloch-Wehba, H., Balkin, J., Berman, E., Crootof, R., Daskal, J., Eichensehr, K., Estrin, M., Haupt, C., Keller, D., Kirtley, J., Kitrosser, H., Klonick, K., Reed, C.: *GLOBAL PLATFORM GOVERNANCE: PRIVATE POWER IN THE SHADOW OF THE STATE*. (2019). https://scholar.smu.edu/smulr/vol72/iss1/9http://digitalrepository.smu.edu
7. Castellanos, M., Dayal, U., Rundensteiner, E.A.: *Enabling Real-Time Business Intelligence*. (2012)
8. Driessen, S.W., Monsieur, G., Van Den Heuvel, W.J.: Data Market Design: A Systematic Literature Review. In: IEEE Access, vol. 10, pp. 33123–33153. Institute of Electrical and Electronics Engineers Inc (2022). https://doi.org/10.1109/ACCESS.2022.3161478
9. Ellili, N., Nobanee, H., Alsaiari, L., Shanti, H., Hillebrand, B., Hassanain, N., Elfout, L.: The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature. J. Finance Data Sci. **9**, 100102 (2023). https://doi.org/10.1016/J.JFDS.2023.100102
10. Elsevier: *Scopus. Content Coverage Guide*. (2023). https://assets.ctfassets.net/o78em1y1w4i4/EX1iy8VxBeQKf8aN2XzOp/c36f79db25484cb38a5972ad9a5472ec/Scopus_ContentCoverage_Guide_WEB.p
11. Engels, B.: Data Governance as the Enabler of the Data Economy. Intereconomics. **54**(4), 216–222 (2019). https://doi.org/10.1007/s10272-019-0827-y
12. Goel, V.: *India's Restaurants Rebel Against Food Delivery Apps*. The New York Times. (2019)., August https://www.nytimes.com/2019/08/29/technology/india-restaurants-logout-delivery-zomato.html
13. GOV.UK. *UK National Data Strategy*. (2020)., December https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy
14. Hao, J., Deng, Z., Li, J.: The evolution of data pricing: From economics to computational intelligence. Heliyon. **9**(9), e20274 (2023). https://doi.org/10.1016/j.heliyon.2023.e20274
15. Harracá, M., Gawer, A.: (n.d.). *Two-faced governance in platform ecosystems*. https://ssrn.com/abstract=4483039
16. IMD: *IMD World Digital Competitiveness Ranking 2023 World Competitiveness Center*. (2023). https://www.imd.org/wp-content/uploads/2023/12/Digital_2023.pdf
17. Janssen, M., Brous, P., Estevez, E., Barbosa, L.S., Janowski, T.: Data governance: Organizing data for trustworthy Artificial Intelligence. Government Inform. Q. **37**(3) (2020). https://doi.org/10.1016/j.giq.2020.101493
18. Koetzle, L., Belissent, J., Peyret, H.: *Design data governance for the data economy*. (2018). https://www.forrester.com/report/Design-Data-Governance-For-The-Data-Economy/RES138595
19. Ladley, J.: Sustaining EIM and the Organization. *Making EIM Enterprise Information Management Work for Business*, 113–118. (2010). https://doi.org/10.1016/B978-0-12-375695-4.00012-6
20. Ladley, J.: Data literacy and concepts. Data Gov. 15–31 (2020a). https://doi.org/10.1016/B978-0-12-815831-9.00003-5
21. Ladley, J.: Prologue: An executive overview. In *Data Governance (Seconds Edition)* (pp. 1–6). Academic Press. (2020b). https://doi.org/10.1016/B978-0-12-815831-9.00001-1
22. Ladley, J.: The data governance business case. Data Gov. 51–60 (2020c). https://doi.org/10.1016/B978-0-12-815831-9.00005-9
23. Muschalle, A., Stahl, F., Loser, A., Vossen, G.: In: Castellanos, M., Dayal, U., Rundensteiner, E.A. (eds.) Pricing Approaches for Data Markets. Springer (2012)
24. OECD: *MEASURING THE VALUE OF DATA AND DATA FLOWS. OECD DIGITAL ECONOMY PAPERS*. (2022).
25. Qiao, Y., Porter, A.L., Huang, Y., Xu, H., Wang, X.: Comparing examiner citations and applicant citations: insights into technology evolution. Scientometrics. (2025). https://doi.org/10.1007/s11192-025-05245-x
26. Ransbotham, S., Kiron, D.: Analytics as a source of business innovation. *MIT Sloan Management Review*, *19*. (2017). https://sloanreview.mit.edu/projects/analytics-as-a-source-of-business-innovation/
27. Revyakin, S.A.: Personal privacy VS. public safety: A hybrid model of the use of smart city solutions in fighting the COVID-19 pandemic in Moscow. Public. Adm. Dev. **42**(5), 281–292 (2022). https://doi.org/10.1002/PAD.1997
28. Singh, P.J., Gurumurthy, A.: *Economic Governance of Data. Balancing individualist-property approaches with a community rights framework*. (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767
29. Soley, A.: *Data Ownership in the Age of the Connected Vehicle*. RTInsights. (2018)., May https://www.rtinsights.com/data-ownership-in-the-age-of-the-connected-vehicle/#:~:text=Drivers%20should%20be%20aware%20of,the%20OEM%2C%20not%20the%20driver
30. Stahl, F., Schomm, F., Vossen, G., Vomfell, L.: A classification framework for data marketplaces. Vietnam J. Comput. Sci. **3**(3), 137–143 (2016). https://doi.org/10.1007/s40595-016-0064-2
31. Treverton, G.F., Esfandiari, P.: *Data: Governance and Geopolitics*. CSIS. (2021)., January https://www.csis.org/analysis/data-governance-and-geopolitics
32. UNCTAD: *VALUE CREATION AND CAPTURE: IMPLICATIONS FOR DEVELOPING COUNTRIES*. (2019)
33. UNCTAD: *Competition law, policy and regulation in the digital era*. (2021). https://unctad.org/system/files/official-document/ciclpd57_en.pdf
34. Vial, G.: Data governance and digital innovation: A translational account of practitioner issues for IS research. Inf. Organ. **33**(1) (2023). https://doi.org/10.1016/j.infoandorg.2023.100450
35. World Economic Forum: *Share to Gain: Unlocking Data Value in Manufacturing*. (2020). https://www3.weforum.org/docs/WEF_Share_to_Gain_Report.pdf

Footnotes

1. https://www.cigionline.org/the-four-domains-of-global-platform-governance/
2. Coverage note: The dataset includes all records retrieved by the specified Scopus query that were assigned to publication years 2012–2024 and were available on the search date. Records assigned to 2024 include online-first and article-in-press items indexed by Scopus at that time; therefore, 2024 does not represent a complete annual cohort. For SciVal-based indicators, the reporting periods differ because of database availability windows and are stated in the respective tables and figures (2012–2023 for the Scholarly output cited in Patents and 2012–2023 for the Policy Impact Indicators (due to the technical restriction of SciVal database), 2018–2022/24 for Tables 5–7, 10–13 and Fig. 7)..
3. The number of joint publications – the intensity of collaboration
4. Intergoverenmental organizations were excluded from the countries list
5. WIPO and European Patent Office were excluded from the countries patent office list

Figures

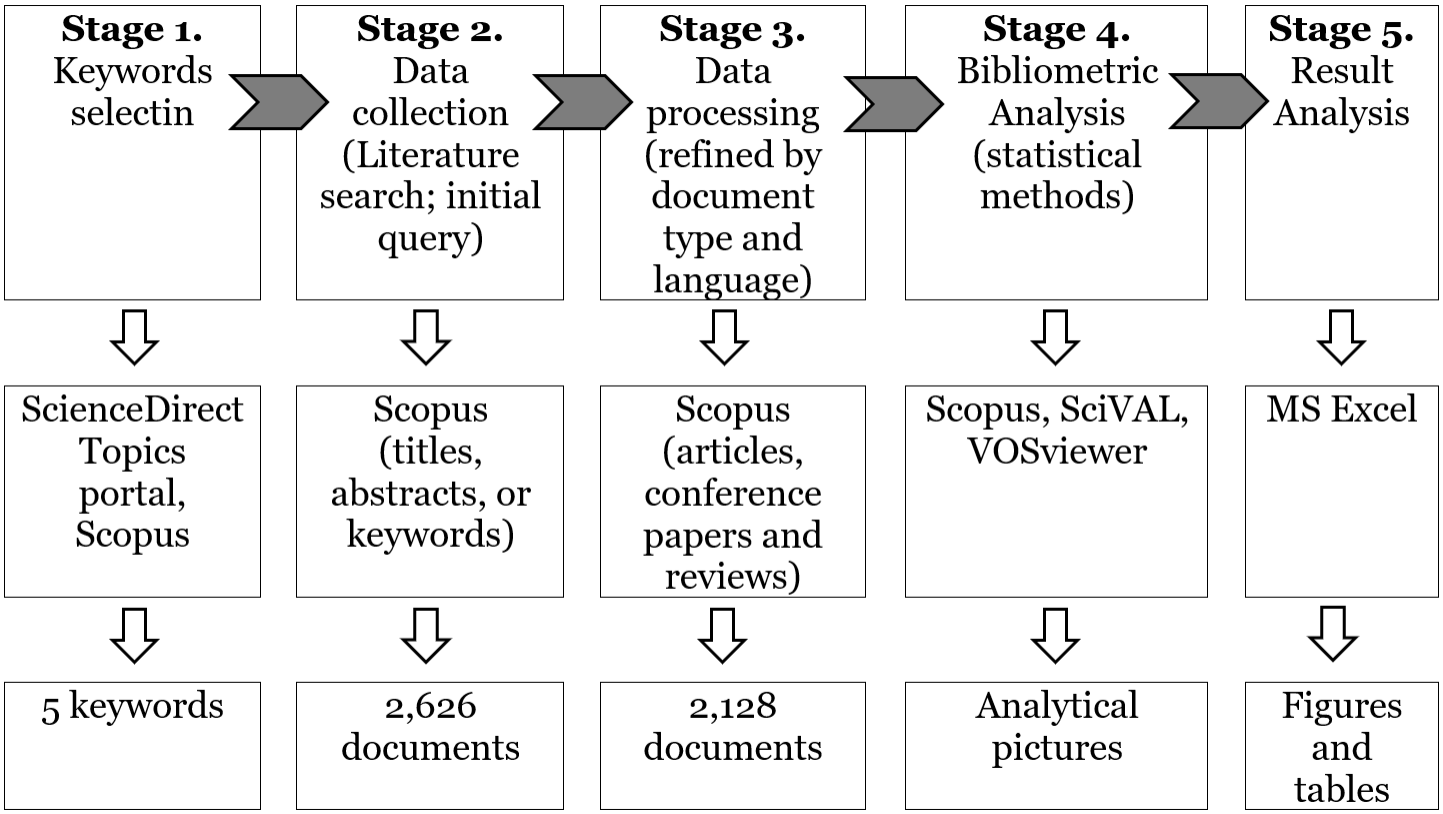

Figure 1

The process applied for dataset preparation on the Data Governance topic.

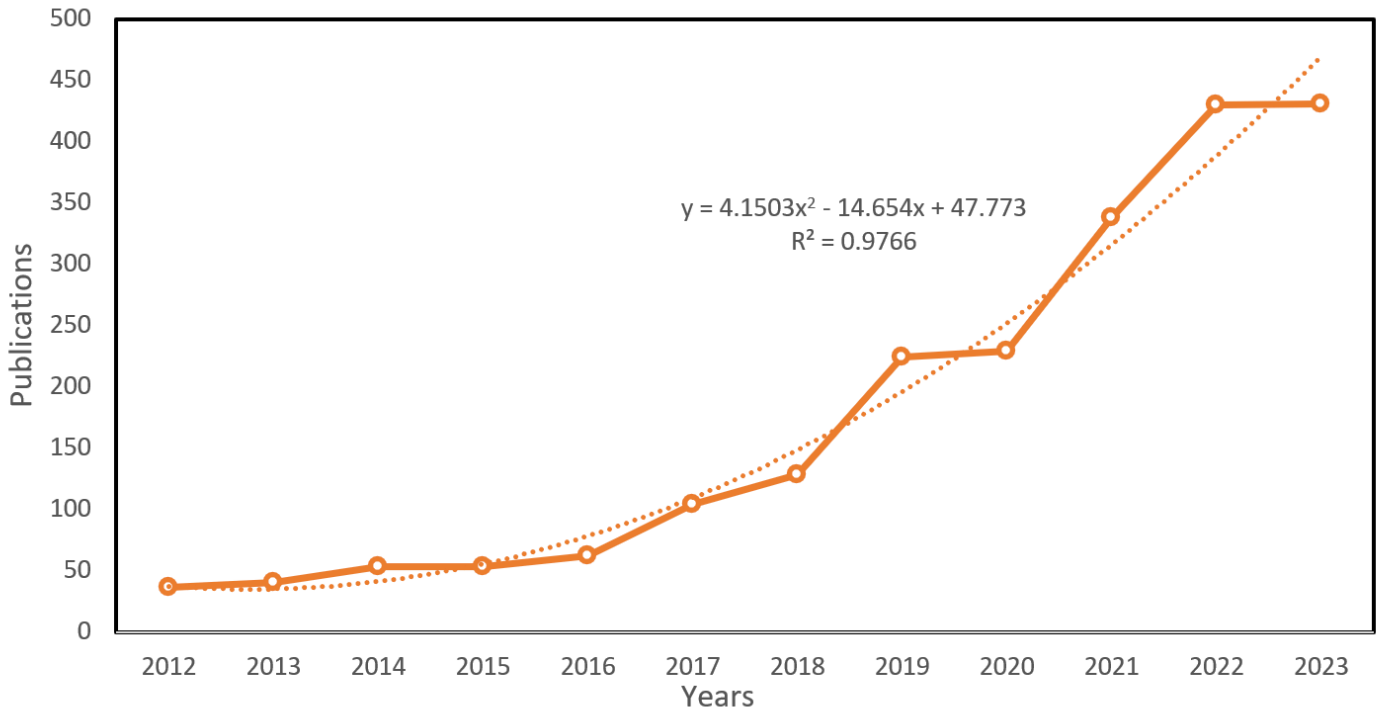

Figure 2

Number of publications per year from 2012 to 2024. (Source: Scopus)

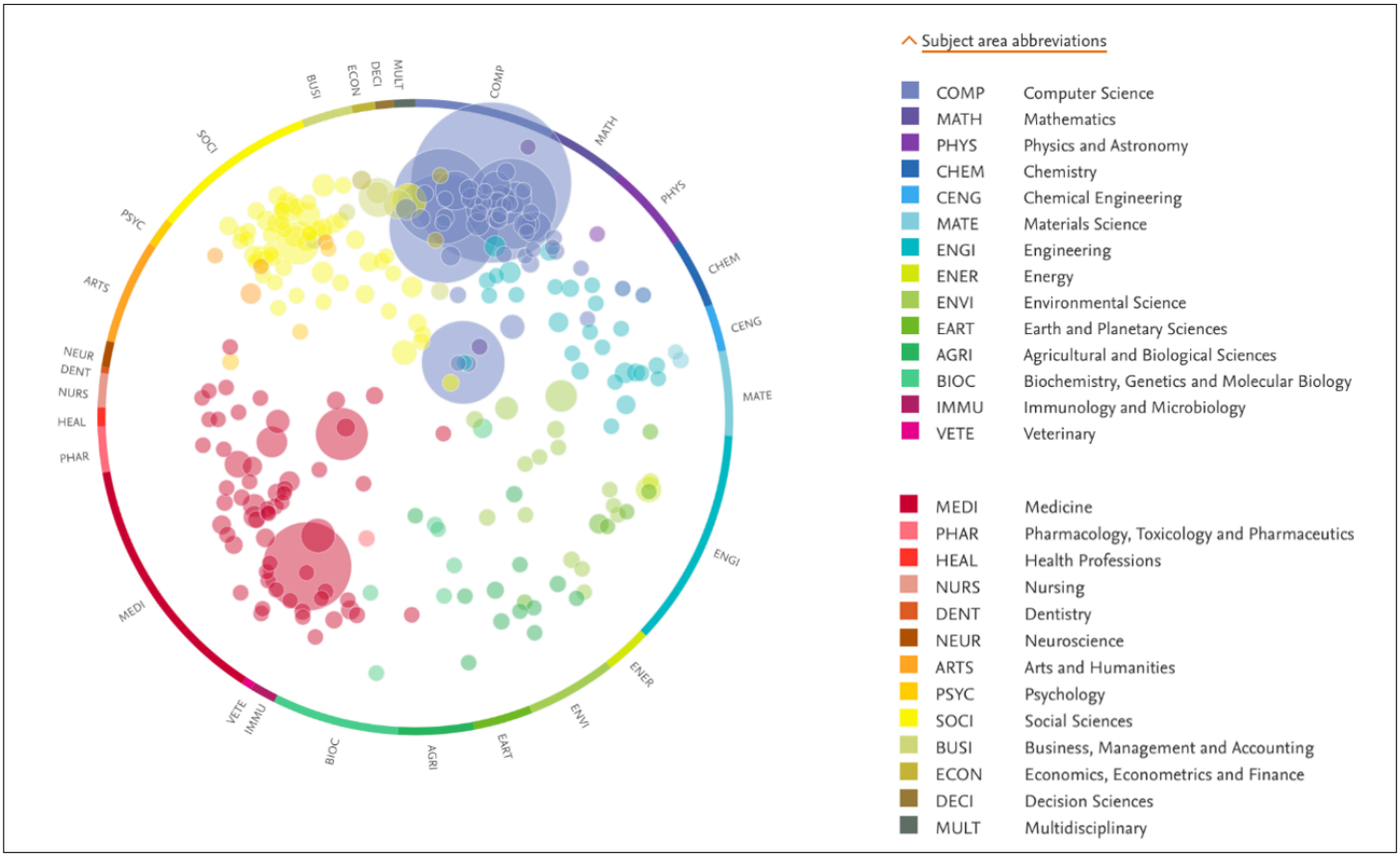

Figure 3

Documents by topic clusters (2018-2024). Source: SciVal)

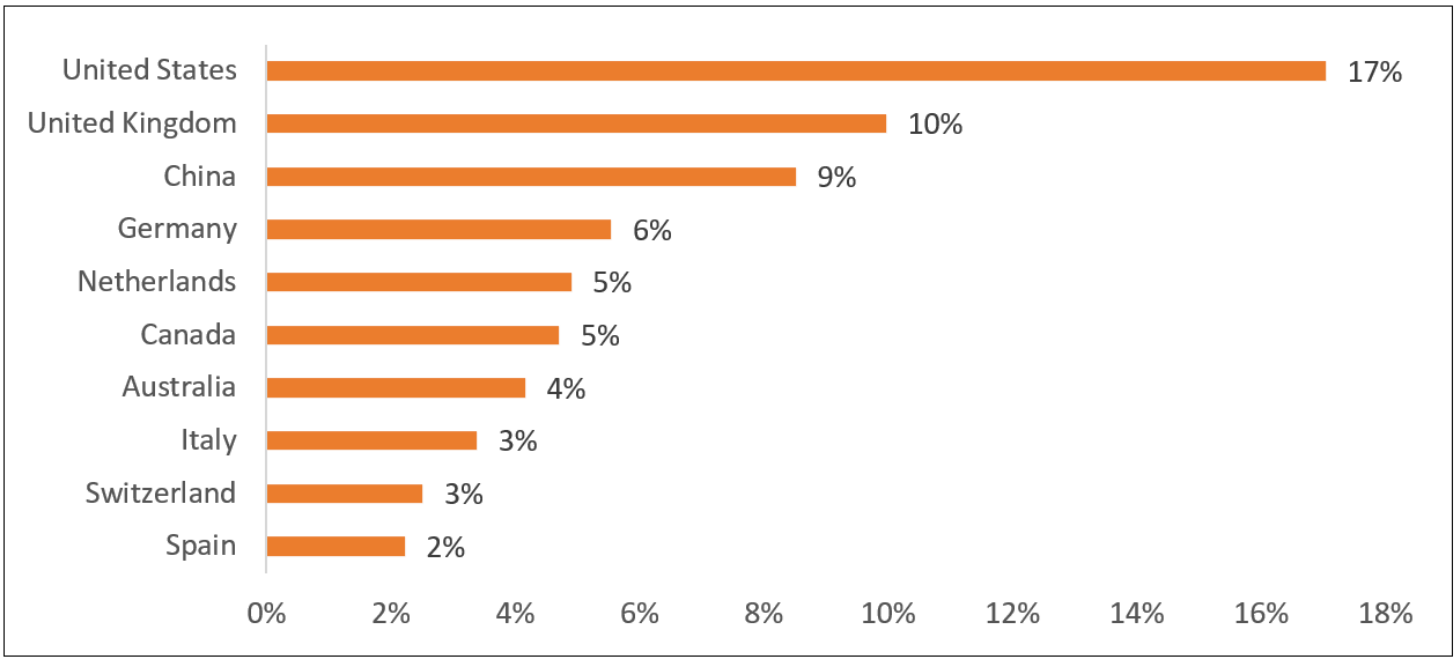

Figure 4

Top-10 countries by publication activity from 2012 to 2024. (Source: Scopus)

Note: The article could have two or more affiliation countries

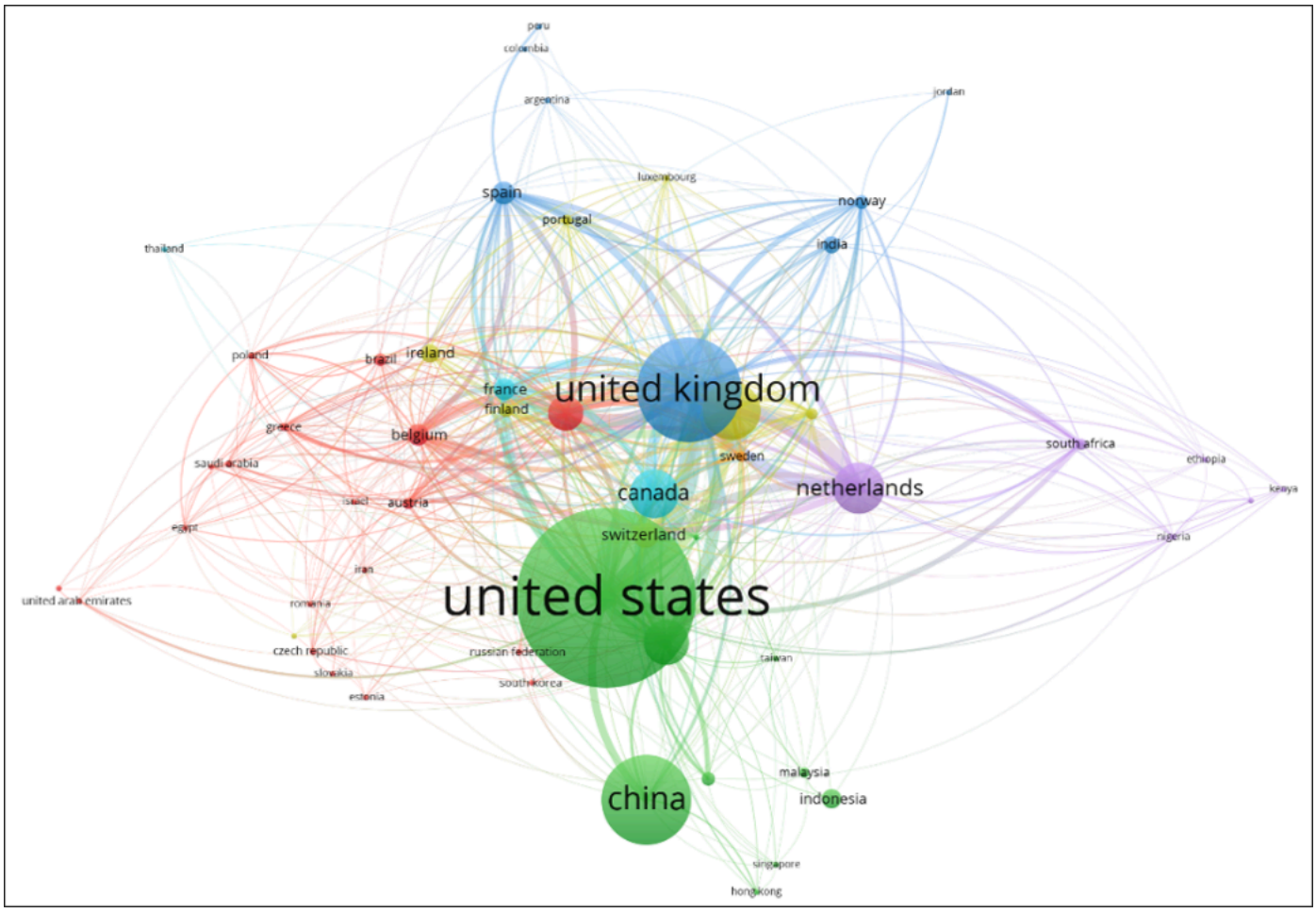

Figure 5

Collaborative network of highly productive countries for data governance research dataset (Source: VOSviewer on Scopus data)

Figure 6

The key topics of strategic documents on data governance adopted by some national governments

Source: (Andreev et al., 2023)

Note. Built on an array of strategic documents in the field of data economics (strategies, programs, plans, policies, etc.) for 2018–2023 by country, including the European Union, UK, Germany, USA, China, India, Singapore, South Korea, Japan, Australia, Canada, Spain, etc., using a national big data analytics system.