IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 August 2024, accepted 30 September 2024, date of publication 8 October 2024, date of current version 22 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3476373

RESEARCH ARTICLE

A Systematic Review of Recent Literature on Data Governance (2017–2023)

KAROL BLIŽNÁK[1], **MICHAL MUNK**[2,3], **AND ANNA PILKOVÁ**[1]
[1]Faculty of Management, Comenius University in Bratislava, 820 05 Bratislava, Slovakia
[2]Science and Research Centre, Faculty of Economics and Administration, University of Pardubice, 532 10 Pardubice, Czech Republic
[3]Department of Computer Science, Constantine the Philosopher University in Nitra, 949 74 Nitra, Slovakia

Corresponding author: Karol Bližnák (bliznak1@uniba.sk)

This work was supported in part by the Scientific Research Project of the Czech Sciences Foundation under Grant 22-22586S, and in part by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences (SAS) under Contract VEGA-1/0734/24.

ABSTRACT In today's rapidly changing environment, organizations are fighting a decisive battle for the most effective use of data. Owing to technological innovation, the volume, velocity, variety, variability, and veracity of data gathered, stored, and processed by organizations in electronic systems are rapidly growing. Analytics, process mining, and artificial intelligence are among the modern application domains of data, enabling data-driven decision making and process innovation for an operating advantage. Data governance, encompassing standards, policies, responsibilities, and relations for managing data, is essential for organizations to maximize the value of the use of data in an effective, cost-efficient, safe, and compliant way. Although data governance has matured as a scientific and business discipline in recent years, the formal definition of data governance and its implementation practices in organizations are still facing ambiguity. New regulations in data protection (e.g., the European Union's General Data Protection Regulation) and safe and ethical data processing (e.g., the European Union's Artificial Intelligence Act) further increase the pressure for compliance and conformity in organizations' management of their data assets. Applying the systematic literature review approach, our objective was to capture state-of-the-art data governance research. The literature review provides an incremental analysis of the most relevant published work on data governance in the period from 2017 to 2023, complementing and enhancing previous systematic literature reviews. The study examines in detail 38 publications, refreshing scientific knowledge and providing further orientation for a growing community of scholars and practitioners in the dynamically evolving data governance discipline.

INDEX TERMS Analytics, artificial intelligence, big data, data governance, systematic literature review.

I. INTRODUCTION

After two decades of maturation, the definition of data governance in scholarly literature remains ambiguous [1], [2], [3], [4], [5], [6], with some convergence and consensus emerging. The authors in [7] synthesized recent scientific knowledge on governance in managing data assets, defining data governance as a specification of "decision rights and accountabilities for an organization's decision-making about its data." Other scholars have also confirmed the role of data governance in formalizing data policies, standards, procedures, accountabilities, and monitoring compliance for the use of data [6], [7], [8], [9], [10], [11].

In addition, we observed the impact of practitioners' publications and institutional standards on shaping the scientific opinion on data governance. The notion of rights, accountabilities and procedures in data work is embedded in the definition data governance by the Data Governance Institute (DGI) [12] as "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

The associate editor coordinating the review of this manuscript and approving it for publication was Huaqing Li.

what circumstances, using what methods." This definition is frequently referred to in recent scholar literatures [1], [2], [3], [13], [14], [15], [16], and [17].

Another stream of thought notes the emancipation of data governance from the data management discipline, which was initially considered to be interchangeable. As a self-contained function, data governance currently encompasses decisions on managing the data asset, whereas data management involves the implementation of these decisions, as described by the study in [4], as seen in [18] and [19]. The Data Management Association (DAMA) practitioners' definition [20], frequently referred to in recent scientific literatures [1], [2], [4], [7], [13], [16], [17], [21], [22], [23], [24], [25], and [26], postulates data governance as: "the exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets." This definition places data governance as a planning/control overlay over data management [2], [3], [6], [22], [27], [28], [29], acknowledging it as a sub-function of corporate/organizational governance [21], [24]. In this context, data governance plays a pivotal role in helping organizations address new regulations in data protection (e.g., the European Union's General Data Protection Regulation (GDPR) [30]) and safe and ethical data processing (e.g., the European Union's Artificial Intelligence Act [31]). Researchers have postulated foundational models to explain the essential organizational and data processing requirements [32], [33] that impact post-regulation data governance practices.

The position of data governance as a compound layer between data management and corporate/organizational governance is also confirmed by the institutional definition of data governance by the International Organization for Standardization (ISO) [34], [35], [36]. The ISO/IEC 38505 standard [34] specifies the governance of data "as a subset or domain of the governance of IT, which itself is a subset or domain of organizational, or in the case of a corporation, corporate governance." The standard provides guidelines for the application of data management as set out by the pivotal ISO/IEC 38500 [37] standard on data management to maximize investment in data use, manage associated risks, and contribute to overall good organizational governance. Well-governed data are a prerequisite for many application domains that provide advantages and deliver value to organizations, including artificial intelligence (AI) [10] and process mining [38]. Hence, the new ISO/IEC 42001 standard for managing artificial intelligence technology also refers to data management and governance as the key constituents [39]. ISO standards are also referred to in recent scientific literature on data governance [1], [6], [16], [17], [24], [25], [40], contributing to the further standardization of the term among scholars.

These generally accepted academic and practitioner definitions delimit the field of study for our research, with the objective of providing insight into the current state of scientific research in data governance as defined above, by systematically locating, analyzing, and evaluating the most recent relevant scientific publications.

The paper is structured as follows. Section II presents recent pivotal literature reviews on data governance, composing the conceptual knowledge base for our study. These reviews have served as the starting point for the analysis of key incremental contributions since their publication. Section III describes the research question, research gap and repeatable procedure we applied to our systematic literature review to answer the research question. The review results and synthesized learning are outlined in Section IV in detail. The conclusion section discusses the research findings and their practical applications as well as directions for future research.

II. CONCEPTUAL BACKGROUND

With our main focus on extracting the state-of-the-art on data governance from recent scientific literature, we are able to build upon existing pivotal literature reviews. Previous literature analyses have introduced and summarized data governance since its introduction, documenting the discipline's partly amorphic evolution and solidifying terms and definitions. These studies also note the gradual exculpation of data governance from data management with its concepts of data quality/data privacy/data security, with which data governance was initially considered interchangeable. The current body of research on data governance reveals a significant gap in establishing a clear consensus on its definition, particularly due to semantic overlaps with related concepts, such as information governance, analytics governance, and data management. The earlier literature reviews indicate the necessity for further work on refining and solidifying a comprehensive data governance model and associated maturity evaluation framework. Additionally, there is a pressing need for a rigorous scientific evaluation of data governance strategies, including a detailed investigation into the motivators, critical success factors, and challenges that organizations face during implementation. A detailed evaluation of the areas for future research is provided in Section V.

The systematic literature review by Al-Ruithe et al. [3] in 2019 has been one of the most influential recent publications on data governance. This study evaluated relevant published works on data governance between 2000-2017. Hence, retrospect represents an important stepstone, summarizing the evolution of data governance since the 1990s in scientific and practice-oriented publications, the discipline's emancipation from pure information technology tasks, the initial ambiguity, and gradual maturation of the terms and definitions. The output of the literature review also included a synthesis of critical success factors and barriers to the implementation of data governance. Our study, which covers publications since 2017, chronologically interlocks with this foundational work, continuously extending the research as a delta review. Our ambition was to include the specifics of recent data governance developments, including new concepts and application areas. Moreover, the related ISO standards [34], [35],

[36] released in 2017 marked a substantial contribution to solidifying the data governance term among both scholars and practitioners, which could not be referred to in [3].

Alhassan et al. [41] published their comprehensive literature review in 2018. They evaluated both academic and practice-oriented literature, arriving at a synthesis of a generic data governance activity model.

The structured literature review by Abraham et al. [7] represents another pivotal study, capturing the state-of-the-art on data governance between 2001-2019. The study proposed a conceptual decomposition of key data governance dimensions (including data governance mechanisms and organizational/data/domain scopes), formulating a set of antecedents with an impact on data governance and the resulting consequences.

Overall, the existing literature reviews refer to a broad range of individual aspects of data governance. Lillie and Eybers [24] published in 2019 a systematic literature review on data governance/data management and the required agile capabilities, with an emphasis on the needs of African organizations. Yebenes and Zorrilla [16] conducted in 2019 a systematic literature review on data governance for Industry 4.0. Langdon and Sikora [42] used a systematic literature review approach to design the concept of a data factory for data products. In 2021, Enders [43] published a literature review on the value of the data. As part of their conceptual research, Lis and Otto [11] conducted a structured literature review evaluating ecosystem data governance.

The most recent literature review also includes the work of Walsh et al. [44], which features organizational motivators for assessing data governance effectiveness. Bassi and Alves-Souza [5] discussed in 2023 a sample of case studies to identify the most impactful challenges for implementing data governance. The literature review by Chandra et al. [45] identified the individual technological elements for effective data governance implementation/utilization based on findings from previous studies. A systematic literature review by Schneider et al. [25] extended the comprehensive data governance concept proposed in [7] into the domain of AI governance.

These literature reviews represent a valuable baseline and roadmap for this study. The research objectives of our study are to advance the state-of-the-art in data governance by conducting a comprehensive review of the relevant literature, and anticipating, capturing and evaluating incremental contributions in the following key areas:

- The development and analysis of the conceptual frameworks, models, and principles that establish data governance as a distinct discipline.
- Identification and examination of critical success factors and challenges in the practical implementation of data governance.
- The evaluation and refinement of models for assessing data governance maturity.
- Exploration of domain-specific approaches to tailoring data governance practices for various application contexts (e.g., AI, analytics, process mining, and Industry 4.0), data (e.g., big data), and deployment (e.g., cloud).

Through this study, we aim to deepen the understanding of these fundamental aspects of data governance, thereby laying the groundwork for future research and practical advancements in the field.

III. RESEARCH METHODOLOGY

For a reproducible study procedure, we applied a systematic literature review approach, as described in [46] and [47]. We scheduled the following steps for our systematic review.

1) Definition of the review's objectives
2) Determination of the exclusion/inclusion criteria
3) Search for studies in digital databases via a keyword-based online query
4) Abstract-based quality/relevance check, removal of duplicates
5) Forward/backward search to increase the coverage of relevant studies in the selected sample
6) Detailed content review, categorization, and description of the studies included in the selected sample, and derivation of quantitative literature statistics.
7) Synthesis of research findings
8) Formulation of conclusions

We articulated the research question for our study as follows: What is the state-of-the-art of scientific research in data governance as a generic discipline shaping the evolution of the term/definition of data governance since the last relevant literature review?

The structured refresh of the status of scientific knowledge in data governance helps fill the research gap: Which notions, frameworks, or definitions have not been covered in recent literature reviews? What has been published by the scholarly community on data governance since recent relevant literature reviews?

We used the Web of Science and Scopus citation databases for the initial online queries of the sample literature. For the subsequent forward/backward search and cross-referencing, we used Google Scholar. The full-text versions were retrieved from the digital libraries of IEEE Xplore, ScienceDirect (Elsevier), SpringerLink, and Emerald.

We used Zotero as a reference manager to organize the necessary information on the bibliography, references, and content through the individual phases of our structured literature review. The captured details included the authors' name, article title, conference/journal name/book title, publication year, publication type, abstract, keywords, and page numbers of the study, and the attachment of the respective article's full-text version where available.

For the keyword-based online queries, we configured an advanced search string containing the main search term, additional terms, and other conditions according to our inclusion/exclusion criteria (publication year, language, and document type), as shown in Table 1.

TABLE 1. Study inclusion and exclusion criteria of systematic review.

Inclusion Criteria	Exclusion Criteria
Publication year 2017-1023	Other publication years
Publications relevant for the research question	Irrelevant publications, e.g., where "data governance" was included in the title/abstract/keywords, but the paper referred to a different research topic
Publications with the main focus on enhancing the scientific knowledge on data governance as a generic notion	Publications related to data governance, but having a different main research subject prevailing over the generic notion of data governance (e.g., a specific sector/subject area, line-of-business, country/region, individual decision domain of data governance, narrow case studies, studies with claims by vendors, etc.)
Peer-reviewed published articles, book chapters and conference proceedings	Not peer-reviewed, pre-prints, other document types
Online availability	Not available online
Published in English	Published in another language
Academic sources only	Practitioners' literature, vendor publications, institutional publications, analyst papers

To address potential biases that could affect the validity and comprehensiveness of the findings, we outline in Section V strategies for future research to minimize the impact of bias, including cross-referencing findings with non-English studies.

We searched for ''data governance'' as the main search term in the publication's title, abstract, and keywords. The term was inserted in double quotation marks (loose phrase search approach) to find studies in which the term appeared together in the searched fields.

Some authors refer to the notion of information governance as an interchangeable term with data governance or a different discipline simultaneously [7], [21]. Given our goal of extracting state-of-the-art generic knowledge on data governance in particular, we did not include ''information governance'' as an explicit main keyword.

We complemented the main term by additional search terms combined by ''OR'' operator to narrow down the search results as shown in Table 2. By including additional search terms denoting selected application/data/deployment domains of data governance, as seen in [10], [15], [16], [21], [23], [25], [27], [28], [29], [34], [40], [42], [43], [45], [48], [49], and [50], we increased the coverage of relevant studies contributing to a broadly applicable definition of data governance. We included the following topics as additional terms in the search query: big data, data lake, data warehouse, management information system, data-driven innovation, analytics, machine learning, artificial intelligence, cloud, large language model, generative AI, decision support, and Internet-of-Things, in their original form, but also using other frequent forms or abbreviations.

Some scholars consider cloud data governance to be a disparate discipline [13], [51] or use ''cloud'' and ''non-cloud'' as a pivotal classification dimension in their literature review [3]. It may be argued that in organizational practice, the cloud represents merely a technical deployment option for data-related processes: whether an organization consumes data storage/processing as a service from a cloud provider or utilizes an in-house IT system for this. We believe that fundamental data governance principles and accountabilities are, to a large extent, generic for both cloud and non-cloud deployment options. In the cloud, the cloud provider receives delegated responsibility for some data governance concepts; however, the concepts remain the same. For this reason, we did not include the term ''cloud'' as a main classification dimension but rather an additional search term.

Whereas practice-oriented sources might provide good guidance for organizational decisions and practical implementation, and some literature reviews in data governance explicitly include them [3], [41], the research value might be possibly diluted by the practical orientation or a specific institutional/analyst focus or vendor claims. In line with our research question, we included only academic contributions in our systematic literature review.

TABLE 2. Advanced search query strings in citation databases.

Citation Database	Query String
Scopus	TITLE-ABS-KEY ("data governance") AND TITLE-ABS-KEY ("big data" OR "data lake" OR " data warehouse" OR "management information system" OR "data-driven innovation" OR "analytics" OR "ml" OR "machine learning" OR "ai" OR "artificial intelligence" OR "cloud" OR "llm" OR "large language model" OR "genai" OR "generative ai" OR "decision support" OR "iot" OR "internet-of-things" OR "internet of things") AND (LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2023)) AND (LIMIT-TO (LANGUAGE , "English"))
Web of Science	TS=("data governance") AND TS=("big data" OR "data lake" OR " data warehouse" OR "management information system" OR "data-driven innovation" OR "analytics" OR "ML" OR "machine learning" OR "AI" OR "Artificial intelligence" OR "cloud" OR "LLM" OR "Large language model" OR "genai" OR "generative AI" OR "decision support" OR "IoT" OR "internet-of-things" OR "internet of things") AND PY=(2017-2023) AND LA=(English)

Note: We manually restricted the results of the Web of Science query to Article, Proceeding Paper, and Book Chapters to keep the query symmetrical with the one used in Scopus.

We sorted the respective results list for each online database by the most cited to facilitate abstract review. Then, we reviewed each abstract to clean the list of results, considering the inclusion and exclusion criteria set for our systematic literature review, as shown in Table 1.

Next, a forward/backward search was applied iteratively to ensure comprehensive coverage of relevant studies and cross-references. In the backward search, we examined the references of papers identified through the initial online search to identify earlier, foundational studies within the study period. Subsequently, we conducted a forward search to identify later studies from the same period that cited these key papers, as they may incorporate more recent scientific insights. External validity, defined in [47] as the generalizability and applicability of findings beyond the scope of the study, was our primary criterion for including additional studies in the review. We added six papers using the forward/backward search that had not been included in the original query results, for example, owing to the featured keywords [4], [5], [7], [27], [41], [44]. Arriving at a pre-selected sample of studies, we retrieved the full-text versions for review and iteratively trimmed the pre-selection again by applying our qualitative relevance criteria.

Although we do not claim that our review is fully exhaustive, the most influential articles were captured and evaluated. The final selected sample of the most relevant published studies that met our inclusion criteria was analyzed and evaluated in detail using full-text versions. We categorized and summarized the eligible publications, derived additional quantitative literature statistics, and combined and synthesized comprehensive research findings. Finally, we present the conclusions and future work, as outlined in Section V.

The individual search and review stages of our systematic literature review yielded the results presented in Table 3.

TABLE 3. Number of studies retrieved per search stage in systematic review.

Search Stage	Number of Studies [a]	
Online keyword-based query	Scopus: 721	Web of Science: 454
Abstract-based relevance check, merging of the results list and removal of duplicates	109	
Forward/backward search	+6	
Detailed review, final sample selection	38	

During the review of the query results (both abstract and full-text), the main reasons for excluding studies from the final selected sample were as follows:

- Not related to data governance
- Related to data governance, but with a different research subject, prevails over the generic notion of data governance. Examples include protecting indigenous rights, obesity research, website data, knowledge creation, corporate performance management, risk management, IT security, and data encryption.
- Duplicates
- Full-text version not available online, implicitly resulting in a lower impact/citation level
- Wrong file format of the full-text version (e.g., slide deck).
- Strong sector/industry focus
- Individual country/regional focus
- Vendor focus
- Individual case studies

We implemented a structured coding scheme to systematically categorize and analyze the data extracted from the literature. Each study was coded according to predefined descriptive and interpretative categories, including approach, approach type, research area, research output type, output, accentuated governance framework dimension, and identified research gaps.

Our analytical framework consisted of three key components as introduced by [47]:

1) Descriptive data synthesis was used to identify trends, relationships, and patterns across studies.
2) A line-of-argument synthesis was used to systematically document and tabulate the approach adopted by each study (see Table 9).
3) A comprehensive systematic review report plotted reference clusters within the foundational model established by the study set (see Table 10).

IV. RESULTS

A. PUBLICATION FREQUENCY

Fig. 1 outlines the annual distribution of the relevant publications within the study period of 2017-2023. The data indicate the steady interest of scholars in data governance research, with peaks in the relevant eligible literature in 2019 (23.68%) and 2023 (21.05%). 2019 witnessed the publication of multiple influential literature reviews summarizing previous studies [3], [7], [16], [24]. We interpret the peak in 2023 as the scholars' response to the onset of AI as a key application domain of data governance [25], [50]. Future literature research will show whether a general upward trend

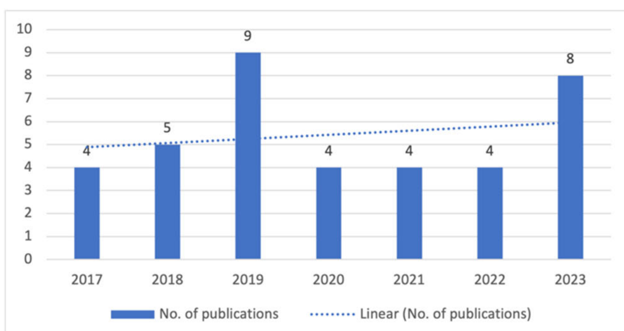

FIGURE 1. Number of relevant publications by year, incl. a trend for the researched period.

in scientific publications in data governance can be confirmed beyond 2023.

B. ANALYSIS OF AUTHORS

We examined the author dimensions in a sample of relevant publications. Table 4 shows the number of contributions by an author or co-author in the analyzed timeframe.

TABLE 4. Number of contributions by author/co-author.

Author	No. of Studies in the Sample	Publications [a]
Al-Ruithe, M.	3	[2], [3], [51]
Benkhelifa, E.	3	[2], [40], [51]
Sammon, D.	3	[4], [41], [44]
Abraham, R.	2	[7], [25]
Alhassan, I.	2	[4], [41]
vom Brocke, J.	2	[7], [25]
Daly, M.	2	[4], [41]
Schneider, J.	2	[7], [25]
Yebenes, J.	2	[16], [29]
Zorrilla, M.	2	[16], [29]

Note: Only authors/co-authors with more than one study in the sample were included.

The author team around Al-Ruithe [2], [3], and [51] can be considered pioneers of structured literature analysis and related conceptual research in data governance. Their impact was also confirmed by the citation statistics within the reviewed literature, as shown in Table 5, where they ranked among the top influencers.

TABLE 5. Top 5 studies with most citations.

Author(s)	No. of Citations	Study [a]
Janssen, Brous, Estevez, Barbosa, Janowski	182	[28]
Abraham, Schneider, vom Brocke	181	[7]
Al-Ruithe, Benkhelifa, Hameed	70	[3]
Alhassan, Sammon, Daly	45	[41]
Al-Badi, Tarhini, Khan	43	[40]

Note: Source: Scopus, retrieved on January 26, 2024.

Janssen, Brous, Estevez, Barbosa, and Janowski are the most frequently cited teams, with impactful conceptual research on the main principles of data governance [28]. In addition, the author team Abraham, Schneider, and vom Brocke, with their structured literature review and conceptual framework research [7], are frequently referred to as sources among the data governance scholar community.

C. PUBLICATION ANALYSIS

We scrutinized the research sample for publication-related attributes including publication outlets and journal names. The graph (Fig. 2) shows the breakdown by publication outlet type (journal articles, conference proceedings, and book sections), with articles (50%) and conference papers (47%) representing the prevailing publication outlet types in the study sample.

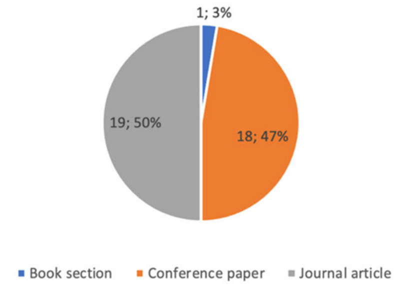

FIGURE 2. Breakdown of studies by publication outlet type.

Next, we examined the journals in which relevant studies from our research sample were published.

TABLE 6. Overview of journals with relevant published studies and corresponding impact factors.

Journal Name	No. of Publications	SCImago Journal Rank 2023
Information and Organization	3	2.01
Information Systems Management	2	1.6
Advances in Artificial Intelligence and Machine Learning	1	0.16
Communications in Computer and Information Science	1	0.2
Computer Standards and Interfaces	1	1.7
Decision Science Letters	1	0.4
Future Internet	1	0.81
Government Information Quarterly	1	2.17
IEEE Access	1	0.96
IEEE Security and Privacy	1	0.79
Information and Management	1	2.59
Information Systems Frontiers	1	1.58
International Journal of Information Management	1	5.78
Journal of Decision Systems	1	0.75
Journal of Enterprise Information Management	1	1.65
Personal and Ubiquitous Computing	1	0.65

Table 6 indicates the broad distribution of scientific journals that publish studies on data governance. Information and Organization (three articles) and Information Systems Management (two articles) were the only journals in which more than one study was included in our sample. Otherwise, the publications were evenly distributed, with one article each. Given that all journals within our sample were indexed by Scopus, we utilized the SCImago Journal Rank (SJR) indicator [52] as a proxy for assessing their impact. SJR reflects the average number of weighted citations received in 2023 by articles published in the preceding three years. Our analysis shows that the journals with the highest volume of data governance studies also tend to have an above-average impact within the sample, indicating significant reach among their respective audiences. The journal scope analysis further revealed that the predominant audience for the data governance studies in our sample comprised scholars and experts in information management and information systems.

D. APPROACH CATEGORIZATION

We applied the approach categorization analysis as described in [53] to provide an overview of the respective scientific approaches used by the authors. Our classification differentiates between empirical and non-empirical approaches. Design science research has also been added as a new empirical category [54].

TABLE 7. Overview of scientific approaches applied by relevant published studies.

Approach Type	Approach	No. of Publications
Empirical	Case study/survey research	2
	Case study	1
	Design science research	1
	Panel study	1
	Survey research	1
Non-empirical	Conceptual research	15
	Literature review/conceptual research	13
	Literature review	4

Table 7 shows that conceptual research (15 studies) and conceptual research combined with a structured literature review (13 studies) were by far the most used approaches to generate and consolidate scientific knowledge in data governance during the research period. The share of empirical research (six studies in total) in the overall sample (38 studies) was still relatively low. Among the limited empirical studies, the panel study by Black et al. [10] provides evidence of individual, organizational, and environmental factors that contribute to successful data governance implementations. The case studies conducted by Knapton [48], Papagiannidis et al. [50], and Zhang et al. [26] provide an empirical understanding of key themes and best-practice recommendations for data governance activities.

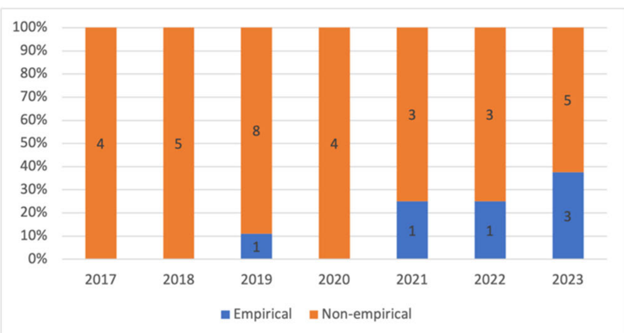

FIGURE 3. Distribution of empirical vs. non-empirical approaches in published studies by year (Percentage and Number of Publications).

While Fig. 3 reveals an increase in the empirical portion in recent years, the studies themselves also suggest more empirical research as a future direction for scientific work in data governance (see also Section V).

E. ANALYSIS OF STATE-OF-THE-ART ON DATA GOVERNANCE

A detailed content examination of the relevant study sample reveals the heterogeneous nature of scientific contributions in the researched timeframe. The authors have been driven by a common interest in consolidating knowledge on data governance by exploring different related research areas. As shown in Table 8, the resulting diversity of research output types contributes to the. state-of-the-art on data governance across the following dimensions:

- Conceptual model/framework/principles/motivators of data governance
- Activity/relationship/workflow model
- Critical success factors/challenges of data governance implementations
- Maturity model for data governance evaluation
- Elaborating data governance concepts specific to the selected application/data/deployment domains AI, big data, analytics, data lakes, clouds, etc.

In some studies, we observed multiple research outputs per study (e.g., a data governance model combined with the respective maturity evaluation framework).

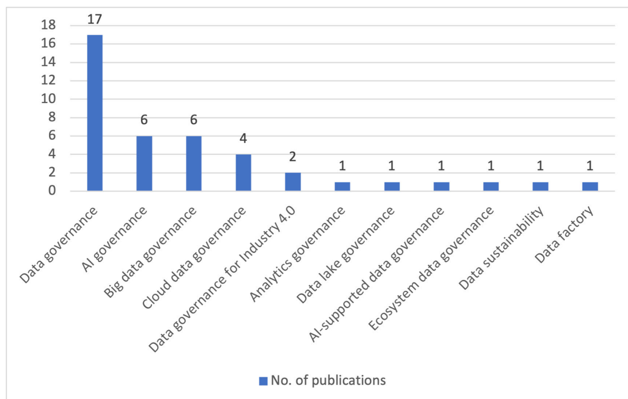

FIGURE 4. Publication frequency by research area in relevant published studies.

The studies at the intersection of the generic research area "data governance" and output type "model/principles" tend to answer best our research question on the state-of-the-art in the generic definition/term of data governance. Several conclusions can be drawn by combining the research area and the output-type data.

- Table 8 confirms that regarding publication frequency, the highest number of relevant studies (10 studies) is situated at the link between "data governance" and "model/principles".
- Fig. 4 reveals that the generic category "data governance" (17 unique studies) is the top value in the

TABLE 8. Overview of relevant published studies with research area and output type.

Research Area	Research Output Type							
	System architecture	Critical success factors	Challenges	Research areas	Activity network	Maturity model	Organizational value of data (governance)	Model/ principles
Data governance	[22]	[4], [10], [45]	[1], [5]	[1], [7], [43]		[4], [6], [44]	[10], [43]	[3], [4], [6], [7], [15], [24], [26], [41], [45], [48]
Big data governance			[55]		[17]	[56]	[9], [23]	[40]
AI governance			[50]					[8], [21], [25], [28], [49], [50]
Analytics governance								[27]
Cloud data governance		[2], [3]	[2], [3]			[13], [51]		[13]
Data lake governance								[57]
Data governance for Industry 4.0						[29]		[16], [29]
AI-supported data governance								[23]
Ecosystem data governance								[11]
Data sustainability								[14]
Data factory								[42]

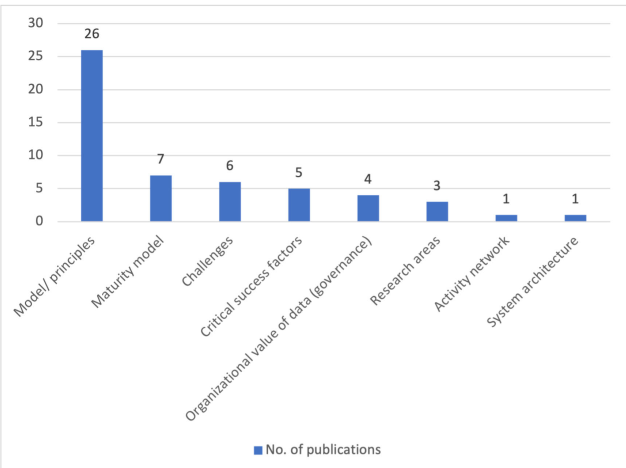

FIGURE 5. Publication frequency in published studies by research output type.

research area dimension, followed by the big data governance data domain (six studies) and the AI governance application domain (six studies).

- Fig. 5 ranks the generic category "model/principles" (26 unique studies) as the top value for the research output type.

The dominance of the studies from the two generic categories "data governance" and "model/principles" as shown in Fig. 4 and 5 confirms that the set-up of our online search supported addressing the research question.

Table 9 provides a chronological overview of the eligible studies, detailing their respective scientific approaches and research outputs.

The research studies have contributed to state-of-the-art conceptual research in data governance. Each published output, including model designs, success factors/challenges, activity models/actor networks, and strategic motivations/principles, as shown in Table 9, helped directly or indirectly enhanced the general definition of data governance. We observed different types of such contributions to the state-of-the-art in the sample, including the following studies:

- Outlining a data governance *conceptual model/ framework*, e.g.,: Al-Ruithe et al. [3] with the conceptual framework for cloud data governance; Abraham et al. [7] with their advanced conceptual framework for data governance, grouping the individual concepts into dimensions and also including the "antecedents" and "consequences" of data governance for a holistic view; Al-Badi et al. [40] synthesizing a governance model for big data; Debattista et al. [23] extending a generic data governance model by the notion of value-driven AI governance; Enders [43] also exploring the data value topic and related research areas; Lillie and Eybers [24] enhancing the generic data governance framework by the notion of agile activities; the factor-based data governance model by Chandra et al. [45] and other generic, conceptual contributions/summaries in [6], [15], [16], and [29], extending a comprehensive data governance conceptual model into the application area of Industry 4.0.
- *Confirming/enhancing an existing model*, for example Schneider et al. [25], referring to and evolving [7] into a generic concept for AI governance.
- Solidifying data governance *semantics/terminology* with their notion of data governance activities [41].
- Introducing an innovative *actor/network* perspective on data governance [17].

TABLE 9. Chronological overview of relevant studies.

Study	Author(s)	Publication Year	Approach	Output
Cloud data governance maturity model	[51]	2017	literature review, conceptual research	Key dimensions and factors for cloud data governance. Cloud data governance maturity model.
Analysis and Classification of Barriers and Critical Success Factors for Implementing a Cloud Data Governance Strategy	[2]	2017	literature review, conceptual research	Classification of success factors and barriers for cloud data governance.
Cloud data governance maturity model	[13]	2017	conceptual research	Cloud data governance model and governance maturity model.
Research on big data governance based on actor-network theory and Petri nets	[17]	2017	conceptual research	Actor/workflow model of (big) data governance.
Exploring big data governance frameworks	[40]	2018	literature review, conceptual research	Big data governance model.
Data governance activities: a comparison between scientific and practice-oriented literature	[41]	2018	literature review, conceptual research	Data governance activity model.
Semantic data ingestion for intelligent, value-driven big data analytics	[23]	2018	conceptual research	Conceptual model for value-driven AI data governance.
Exploring the Value of Data – A Research Agenda	[43]	2018	literature review, conceptual research	Value-based data governance. Research agenda to further explore the value of data.
Research and application of enterprise big data governance	[15]	2018	conceptual research	Data governance model.
Big data governance, dynamic capability and decision-making effectiveness: Fuzzy sets approach	[9]	2019	survey research	Proven causal relationship between big data governance and effective decision making.
Data governance: A conceptual framework, structured review, and research agenda	[7]	2019	literature review, conceptual research	Data governance model. Identified gaps for further research.
Evaluation of big data governance - Combining a multi-criteria approach and systems theory	[56]	2019	conceptual research	Big data governance maturity model.
A systematic literature review of data governance and cloud data governance	[3]	2019	literature review, conceptual research	Critical success factors/barriers for implementation of data governance. Key dimensions of data governance. Challenges of implementing cloud data governance.
Critical Success Factors for Data Governance: A Theory Building Approach	[4]	2019	literature review, conceptual research	Critical success factors for data governance, including an action-based data governance model. Data governance maturity evaluation matrix.
Overview of Data Governance in Business Contexts	[6]	2019	conceptual research	Data governance model and maturity model.
Data Science Data Governance [AI Ethics]	[49]	2019	conceptual research	Responsible data science practices for AI ethics.
Identifying the constructs and agile capabilities of data governance and data management: A review of the literature	[24]	2019	literature review	Agile capabilities for data governance and data management.
Towards a data governance framework for third generation platforms	[16]	2019	literature review, conceptual research	Contribution to a simple, maintainable and extendable data governance model for Industry 4.0.
Data lake governance: Towards a systemic and natural ecosystem analogy	[57]	2020	conceptual research	Data lake governance model based on natural manner and systematic manner as multidisciplinary approaches.
Data governance: Organizing data for trustworthy Artificial Intelligence	[28]	2020	conceptual research	Data governance principles for AI based on big data.
Creating a Data Factory for Data Products	[42]	2020	literature review, conceptual research	Data factory model.
Analysis of data governance implications on big data	[55]	2020	conceptual research	Big data governance dynamics, challenges, opportunities and risks.

TABLE 9. ***(Continued.)*** **Chronological overview of relevant studies.**

Governance of Ethical and Trustworthy AI Systems: Research Gaps in the ECCOLA Method	[21]	2021	conceptual research	Typology combining themes/practices for ethical AI development.
An Ontological-Based Model to Data Governance for Big Data	[22]	2021	literature review, conceptual research	Autonomous system architecture as an ontology-based reasoning distributed system for the decision associated with the data governance processes.
Toward big data and analytics governance: Redefining structural governance mechanisms	[27]	2021	design science research	Roles and responsibilities for data/analytics governance.
Towards a taxonomy of ecosystem data governance	[11]	2021	literature review, conceptual research	Taxonomy of ecosystem data governance.
Towards Data Governance for Federated Machine Learning	[8]	2022	conceptual research	Approach for data governance for federated machine learning.
Grounding data governance motivations: a review of the literature	[44]	2022	literature review, conceptual research	Five decision domains for assessing data governance effectiveness.
Data Matters: A Strategic Action Framework for Data Governance	[26]	2022	case study	Strategic action framework for data governance.
A reference framework for the implementation of data governance systems for industry 4.0	[29]	2022	conceptual research	Data governance model and maturity model for Industry 4.0.
Challenges to Implementing Effective Data Governance: A Literature Review	[5]	2023	literature review	Most influential data governance challenges.
Data governance and the secondary use of data: The board influence	[10]	2023	panel study	Data governance as a critical enabler of competitive advantage.
Control and Data Integrity are Important Factors of Data Governance Technology	[45]	2023	literature review	Data governance model with primary technological factors for data governance efficacy.
Data sustainability: Data governance in data infrastructures across technological and human generations	[14]	2023	conceptual research	Sustainability challenges in data governance.
Exploring Mid-Market Strategies for Big Data Governance	[48]	2023	case study, survey research	Major themes for data governance strategies.
Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes	[50]	2023	case study, survey research	AI governance model. Key challenges and recommended actions.
Artificial Intelligence Governance For Businesses	[25]	2023	literature review, conceptual research	AI governance model.
Data governance and digital innovation: A translational account of practitioner issues for IS research	[1]	2023	conceptual research	Practitioners' challenges in data governance. Research areas in data governance.

- Elaborating the emerging, urgent topic of *AI Ethics* [21], [49].
- Emphasizing selected concepts of the overall data governance model by declaring them *critical success factors* [2], [3], [4] and/or *challenges* [1], [5], [50].
- Proposed *evaluation models for data governance maturity* [4], [6], [13], [29], [44], [51], [56].
- Exploring *key principles, strategies, and motivators* for data governance [26], [28], [48].

Table 10 summarizes the conceptual framework for data governance, as reflected in recent literature on data governance. We used the comprehensive conceptual framework outlined in [7] as the baseline. Where appropriate, we adopted the model by extending the list of the concepts in [7] with additional notions, as referred to by other relevant published work from our systematic literature review. We regard the framework in [7] as the most adequate synthesis at present of the previously available conceptual work, serving as effective guidance for both scholars and practitioners.

The data governance model in [7] consists of six dimensions:

1) Governance mechanisms
2) Organizational scope
3) Data scope
4) Domain scope
5) Antecedents
6) Consequences.

TABLE 10. Overview of state-of-the-art data governance framework dimensions with referencing published studies.

Dimension	Sub-dimension	Concept/Factor/Effect	Referencing Studies
Antecedents: "Contingency factors which impact the adoption and implementation of data governance" [7]	External	Legal and regulatory requirements, market volatility, industry, country	[7], [10], [25]
	Internal	Organizational strategy, IT strategy, diversification, location of decision-making authority, degree of process harmonization/integration, IT architecture, organization culture, data-driven mindset, data strategy, data governance maturity, senior management support, business case for data governance, availability of resources for data governance, business agility, involvement of data stakeholders, business ownership of data analysis	[2], [3], [6], [7], [10], [25], [27], [29], [40], [51]
Governance mechanisms: Formal structures that "help to plan and control data management activities" [7]	Structural	Roles and responsibilities, location of decision-making authority	[1], [2], [3], [4], [7], [13], [16], [24], [25], [27], [28], [29], [40], [45], [50], [51], [56]
	Procedural	Data strategy, data governance strategy/roadmap, policies, standards, processes, procedures, contractual agreements, performance measurement/feedback, cost measurement, compliance monitoring, data governance tools, issue management, change management	[1], [2], [3], [4], [7], [13], [16], [17], [25], [28], [29], [40], [45], [50], [51]
	Relational	Communication, training, coordination of decision making	[1], [2], [3], [7], [13], [21], [25], [26], [40], [45], [50], [51]
Organizational scope: "The unit of analysis" [7]	Intra-organizational	Data governance organization on project/company level	[2], [3], [6], [7], [13], [15], [17], [24], [25], [29], [43], [45], [51]
	Inter-organizational	Ecosystem data governance	[7], [11], [14], [25], [26], [28], [29], [49], [50]
Domain scope: "Main data decision domains" [7]		Data quality, data quantity, data integrity, data standardization, data security, data privacy, data availability, data requirements, data access, data sharing/transfer, data diversity, data architecture, data integration, data lifecycle, meta data, data classification, data transparency, auditability, data value, data storage, infrastructure, tools, technologies, performance, availability in the cloud, service level agreement in the cloud, AI ethics, data usage by AI	[3], [6], [7], [13], [15], [17], [21], [25], [28], [29], [41], [45], [49], [56]
Data scope: Data areas with "partially different requirements on data governance" [7]	Traditional data	Master data, transactional data, reference data, analytical data	[7], [15], [27]
	Big data	Web and social media data, machine-generated data, streaming data, biometric data, consumer data, unstructured data	[6], [7], [15], [16], [28], [29], [40], [56]
Consequences: "The effects of data governance" [7]		Intermediate performance effects, risk/compliance management, explainability of data-driven decisions, lineage/trustworthiness, social impact, environmental impact, orchestration	[2], [6], [7], [10], [21], [25], [26], [29], [43], [49], [50]

Hence, the model provides a logical structure and conceptual foundation for integrating individual data governance concepts, factors, and effects from foundational research and new insights from our systematic literature review. To render the state-of-the art on data governance as the sum of the contributions of the eligible studies, we eventually mapped the framework's dimensions to publications referencing, accentuating, or acknowledging the respective dimension as part of the general model.

In the study texts, we also examined the relative prominence of individual data governance concepts when forming the state-of-the-art framework. The results in Fig. 6 are only approximate, as we had to judge whether the respective term was indeed used in the context of a data governance framework and not just as a passing mention. However, the aggregated results provide a useful overview of the most frequently highlighted individual data governance concepts within research studies.

The decision domain of the data quality is the most frequently mentioned concept. This is partly because the application domains of AI and analytics require high-quality data, thus affecting related data governance decisions. In addition, the historical evolution of data governance plays a role here and its ties with technical data management, where data quality is considered a priority. Policies, standards, procedures, roles, and responsibilities are the key concepts of every data governance implementation. Thus, their strong positions within the rankings are logical. Applying advanced text analysis/natural language processing methods in papers' content/abstracts and/or in database queries related to data governance to examine the relationships between these terms and the evolution of their ranking position

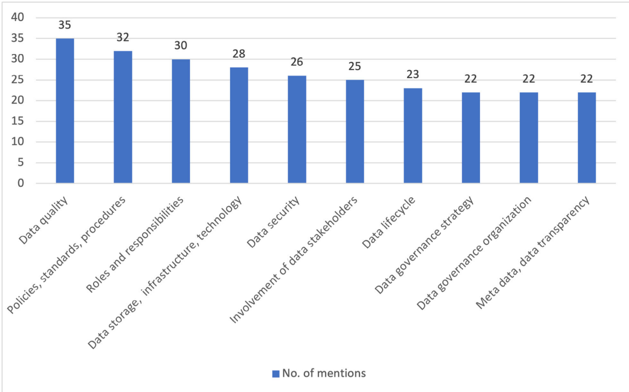

FIGURE 6. Top 10 data governance concepts by the number of referencing relevant published studies.

over time might be an impactful research idea for future work.

V. CONCLUSION AND FUTURE WORKS

A. RESEARCH AGENDA

In the era of AI and big data, progress in scientific research on governance exercised over data management is highly desirable. We synthesized the call for further research in this field, as presented in individual research studies. There is still a gap in the consensus among researchers regarding the definition of data governance, including the semantic overlaps between data governance, information governance, analytics governance, and data management. The reviewed literature suggests future work on solidifying the data governance model and the related maturity evaluation framework. Organizations may also benefit from a thorough scientific evaluation of data governance strategies and motivators, critical success factors, and challenges pertinent to the practical implementation of data governance. Future researchers should also enhance the actor/activity model of data governance, the implications of data governance in an ecosystem beyond the boundaries of one organization, further elaborate the notion of data value, and conduct empirical, case-study based research in the development of data governance implementation over time.

Beyond the core model focus, future works might:

- Apply enhanced scientific methods in new examinations, as well as validate, extend, refine, and generalize the existing knowledge on data governance.
- More frequently use empirical approaches, including case studies, longitudinal/panel data, and focus groups. Further approaches proposed for future work include advanced content analysis methods, relationship analyses, and in-depth analyses.
- Charter the surrounding, emerging topics, including the concept of automated/AI-supported data governance, data factories for data products, or the domains for which generic data governance is specifically applied, including data governance for AI, advanced analytics, process mining, big data, cloud data, and Industry 4.0.
- Explore sector-specific characteristics (e.g., data governance in healthcare, higher education).
- Evaluate the impact of AI ethics on data/AI/analytics governance in light of the recent regulations, focusing particularly on ethical practices in work with the data asset.

B. CONCLUSION

The objective of our study was to perform a systematic literature review, capturing the state-of-the-art on data governance in recently published work. As an important step, we charted previous pivotal literature analyses on data governance. These had been tracking data governance since its introduction, noting the term's evolution and separation from the technical data management discipline, as reflected in the hitherto available pioneering works.

To determine the incremental contribution to the state-of-the-art on data governance, we performed a repeatable, structured search in leading scientific databases. We conducted a thorough analysis of the metadata and content within the retrieved samples of relevant studies. The outcomes yielded several interesting findings for the researched period between 2017-2023:

We observed a healthy publication frequency in the scientific literature on data governance with an ongoing upward trend. The evaluation of the individual authors' publication frequency and impact helped to identify key players in the data governance scholar community.

Analysis of publication outlets revealed a broad distribution of studies relevant to data governance across journals, conferences, and books, targeting the main audience of information systems/information management scholars and practitioners.

Regarding the scientific approaches used in the studies, we confirmed the current prevalence of non-empirical scientific methodologies, with a recent uptake observed in empirical approaches as well.

When analyzing the output types of the relevant studies, we observed good coverage of the input to the conceptual model. This means that recently published research has significantly contributed to solidifying the generic definition and framework of data governance. To provide orientation aid to future researchers, we have included a chronological timeline featuring relevant studies with short summaries. Our synthesis of the state-of-the-art data governance conceptual framework, as proposed and/or enhanced by relevant studies, can serve as a stabilizing foundation for future incremental research in this area.

The limitations of our study include the focused selection of keywords for online literature search, which poses a methodological challenge. Our choice of the single main search term "data governance" was deliberate, excluding the terms of "information governance", "analytics governance", or "data management" for reasons discussed in

Section III. Future research in this area may include a comparative analysis to reveal the extent to which these terms are interchangeable, overlapping, or disjunct. Future research should also chart the chronological evolution of the notion of data governance since its introduction. We are also aware that excluding practitioners' and institutional publications from the scope of our systematic literature research might have led to some omissions in the state-of-the-art. However, the setup of our study was intended to keep the outcome undiluted by the vendor/institutional focus. Future work will involve broader, more comprehensive, systematic literature reviews across scholars and practice-oriented literature as a valuable update. In addition, we acknowledge that not including non-English studies in the scope might have introduced biases affecting the validity and comprehensiveness of the findings. For future research, we suggest developing strategies to minimize language biases and ensure a more balanced and comprehensive review. The strategies will involve including non-English studies in search and cross-referencing, collaborating with experts fluent in different languages, using of translation tools for accurate interpretation of non-English studies, and comparing findings across languages.

We believe that our systematic literature review of state-of-the-art data governance has direct practical implications for both scholars and practitioners in data governance. An overview of recent knowledge will serve as a guide for better orientation in most current data governance concepts, as well as in recent publications. Our intention is to contribute to the further solidification of data governance as a scientific and organizational discipline, and to map out key directions for future research.

REFERENCES

[1] G. Vial, "Data governance and digital innovation: A translational account of practitioner issues for IS research," *Inf. Org.*, vol. 33, no. 1, Mar. 2023, Art. no. 100450.

[2] M. Al-Ruithe and E. Benkhelifa, "Analysis and classification of barriers and critical success factors for implementing a cloud data governance strategy," *Proc. Comput. Sci.*, vol. 113, pp. 223–232, May 2017.

[3] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, "A systematic literature review of data governance and cloud data governance," *Pers. Ubiquitous Comput.*, vol. 23, nos. 5–6, pp. 839–859, Nov. 2019.

[4] I. Alhassan, D. Sammon, and M. Daly, "Critical success factors for data governance: A theory building approach," *Inf. Syst. Manage.*, vol. 36, no. 2, pp. 98–110, Apr. 2019.

[5] C. Bassi and S. Alves-Souza, "Challenges to implementing effective data governance: A literature review," presented at the *Proc. 15th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage.*, Apr. 2023.

[6] L. Jimenez, J. Polo, and N. Duarte, "Overview of data governance in business contexts," *IOP Conference Ser., Materials Science Engineering*, vol. 519, no. 1, 2018, Art. no. 012023.

[7] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *Int. J. Inf. Manage.*, vol. 49, pp. 424–438, Dec. 2019.

[8] J. A. Peregrina, G. Ortiz, and C. Zirpins, "Towards data governance for federated machine learning," in *Communications in Computer and Information Science*, C. Zirpins, G. Ortiz, Z. Nochta, O. Waldhorst, J. Soldani, M. Villari, and D. Tamburri, Eds., 2022, pp. 59–71.

[9] M. A. S. Abbady, M. Akkaya, and A. Sari, "Big data governance, dynamic capability and decision-making effectiveness: Fuzzy sets approach," *Decis. Sci. Lett.*, vol. 8, no. 4, pp. 429–440, 2019.

[10] S. Black, M. Davern, S. B. Maynard, and H. Nasser, "Data governance and the secondary use of data: The board influence," *Inf. Org.*, vol. 33, no. 2, Jun. 2023, Art. no. 100447.

[11] D. Lis and B. Otto, "Towards a taxonomy of ecosystem data governance," presented at the *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, Jun. 2021.

[12] *Definitions of Data Governance*. Accessed: Jan. 10, 2024. [Online]. Available: https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/

[13] G. Cheng, Y. Li, Z. Gao, and X. Liu, "Cloud data governance maturity model," presented at the *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Nov. 2017.

[14] S. L. Jarvenpaa and A. Essén, "Data sustainability: Data governance in data infrastructures across technological and human generations," *Inf. Org.*, vol. 33, no. 1, Mar. 2023, Art. no. 100449.

[15] C. Ke and K. Wang, "Research and application of enterprise big data governance," presented at the *Proc. 2nd Int. Conf. Comput. Sci. Appl. Eng.*, Oct. 2018.

[16] J. Yebenes and M. Zorrilla, "Towards a data governance framework for third generation platforms," *Proc. Comput. Sci.*, vol. 151, pp. 614–621, Aug. 2019.

[17] S. Zhang, H. Gao, L. Yang, and J. Song, "Research on big data governance based on actor-network theory and Petri nets," presented at the *Proc. IEEE 21st Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, Apr. 2017.

[18] X. Fu, A. Wojak, D. Neagu, M. Ridley, and K. Travis, "Data governance in predictive toxicology: A review," *J. Cheminformatics*, vol. 3, no. 1, pp. 1–20, Dec. 2011.

[19] V. Khatri and C. V. Brown, "Designing data governance," *Commun. ACM*, vol. 53, no. 1, pp. 148–152, Jan. 2010.

[20] M. Mosley, M. H. Brackett, S. Earley, and D. Henderson. (2009). *Data Administration Management Association, DAMA Guide To the Data Management Body of Knowledge (DAMA-DMBOK Guide)*. [Online]. Available: https://proquest.safaribooksonline.com/9781935504009

[21] M. Agbese, H.-K. Alanen, J. Antikainen, E. Halme, H. Isomäki, M. Jantunen, K.-K. Kemell, R. Rousi, H. Vainio-Pekka, and V. Vakkuri, "Governance of ethical and trustworthy al systems: Research gaps in the ECCOLA method," presented at the *Proc. IEEE 29th Int. Requirements Eng. Conf. Workshops (REW)*, Sep. 2021.

[22] A. Castro, V. A. Villagrá, P. García, D. Rivera, and D. Toledo, "An ontological-based model to data governance for big data," *IEEE Access*, vol. 9, pp. 109943–109959, 2021.

[23] J. Debattista, J. Attard, and R. Brennan, "Semantic data ingestion for intelligent, value-driven big data analytics," presented at the *Proc. 4th Int. Conf. Big Data Innov. Appl. (Innovate-Data)*, Aug. 2018.

[24] T. Lillie and S. Eybers, "Identifying the constructs and agile capabilities of data governance and data management: A review of the literature," *Commun. Comput. Inf. Sci.*, vol. 933, pp. 313–326, May 2019.

[25] J. Schneider, R. Abraham, C. Meske, and J. Vom Brocke, "Artificial intelligence governance for businesses," *Inf. Syst. Manage.*, vol. 40, no. 3, pp. 229–249, Jul. 2023.

[26] Q. Zhang, X. Sun, and M. Zhang, "Data matters: A strategic action framework for data governance," *Inf. Manage.*, vol. 59, no. 4, Jun. 2022, Art. no. 103642.

[27] M. Fadler and C. Legner, "Toward big data and analytics governance: Redefining structural governance mechanisms," presented at the *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, May 2021.

[28] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy artificial intelligence," *Government Inf. Quart.*, vol. 37, no. 3, Jul. 2020, Art. no. 101493.

[29] M. Zorrilla and J. Yebenes, "A reference framework for the implementation of data governance systems for Industry 4.0," *Comput. Standards Interface*, vol. 81, Apr. 2022, Art. no. 103595.

[30] *GDPR Archives-GDPR.eu*. Accessed: Jan. 10, 2024. [Online]. Available: https://gdpr.eu/tag/gdpr/

[31] (2021). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Accessed: Jan. 10, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX

[32] P. Voigt and A. von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Cham, Switzerland: Springer, 2017.

[33] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Putting AI ethics into practice: The hourglass model of organizational AI governance," 2022, *arXiv:2206.00335*.

[34] *Information Technology—Governance of IT—Governance of Data—Part 1: Application of ISO/IEC 38500 To the Governance of Data*, Standard ISO/IEC 38505-1:2017, 2017.

[35] *Information Technology—Governance of IT—Governance of Data—Part 2: Implications of ISO/IEC 38505-1 for Data Management*, Standard ISO/IEC TR 38505-2:2018, 2018.

[36] *Information Technology—Governance of Data—Part 3: Guidelines for Data Classification*, Standard ISO/IEC TS 38505-3:2021, 2021.

[37] *Information Technology—Governance of IT for the Organization*, Standard ISO/IEC 38500:2015, 2015.

[38] M. Jans, J. D. Weerdt, B. Depaire, M. Dumas, and G. Janssenswillen, "Conformance checking in process mining," *Inf. Syst.*, vol. 102, May 2021, Art. no. 101851.

[39] *Information Technology—Artificial Intelligence—Management System*, Standard ISO/IEC 42001:2023, 2023.

[40] A. Al-Badi, A. Tarhini, and A. I. Khan, "Exploring big data governance frameworks," *Proc. Comput. Sci.*, vol. 141, pp. 271–277, Apr. 2018.

[41] I. Alhassan, D. Sammon, and M. Daly, "Data governance activities: A comparison between scientific and practice-oriented literature," *J. Enterprise Inf. Manage.*, vol. 31, no. 2, pp. 300–316, Mar. 2018.

[42] C. Schlueter Langdon and R. Sikora, "Creating a data factory for data products," presented at the Lecture Notes in Business Information Processing, K. R. Lang, J. Xu, B. Zhu, X. Liu, M. J. Shaw, H. Zhang, and M. Fan, Eds., 2020, pp. 43–55.

[43] T. Enders, "Exploring the value of data—A research agenda," presented at the Lecture Notes in Business Information Processing, G. Satzger, N. Kühl, P. Hottum, M. Zaki, and L. Patrício, Eds., 2018, pp. 274–286.

[44] M. J. Walsh, J. McAvoy, and D. Sammon, "Grounding data governance motivations: A review of the literature," *J. Decis. Syst.*, vol. 31, no. sup1, pp. 282–298, Dec. 2022.

[45] Y. U. Chandra, H. Prabowo, F. L. Gaol, and B. Purwandari, "Control and data integrity are important factors of data governance technology," presented at the *Proc. 10th Int. Conf. ICT Smart Soc. (ICISS)*, Sep. 2023.

[46] D. Tranfield, D. Denyer, and P. Smart, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," *Brit. J. Manage.*, vol. 14, no. 3, pp. 207–222, Sep. 2003.

[47] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," vol. 2, Jan. 2007.

[48] K. Knapton, "Exploring mid-market strategies for big data governance," *Adv. Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 816–838, 2023.

[49] J. A. Kroll, "Data science data governance [AI Ethics]," *IEEE Secur. Privacy*, vol. 16, no. 6, pp. 61–70, Nov. 2018.

[50] E. Papagiannidis, I. M. Enholm, C. Dremel, P. Mikalef, and J. Krogstie, "Toward AI governance: Identifying best practices and potential barriers and outcomes," *Inf. Syst. Frontiers*, vol. 25, no. 1, pp. 123–141, Feb. 2023.

[51] M. Al-Ruithe and E. Benkhelifa, "Cloud data governance maturity model," presented at the *Proc. ACM Int. Conf. Proc. Ser.*, 2017.

[52] *SJR-Help*. Accessed: Aug. 27, 2024. [Online]. Available: https://www.scimagojr.com/help.php

[53] N. Kryvinska, R. Olexova, P. Dohmen, and C. Strauss, "The S-D logic phenomenon-conceptualization and systematization by reviewing the literature of a decade(2004–2013)," *J. Service Sci. Res.*, vol. 5, no. 1, pp. 35–94, Nov. 2013.

[54] R. H. Von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quart.*, vol. 28, no. 1, pp. 75–105, 2004.

[55] L. Trom and J. Cronje, "Analysis of data governance implications on big data," *Advances Information Communication*, vol. 69, pp. 645–654, May 2020.

[56] J. Akoka and I. Comyn-Wattiau, "Evaluation of big data governance–combining a multi-criteria approach and systems theory," presented at the *Proc. IEEE World Congr. Services (SERVICES)*, Jul. 2019.

[57] M. Derakhshannia, C. Gervet, H. Hajj-Hassan, A. Laurent, and A. Martin, "Data lake governance: Towards a systemic and natural ecosystem analogy," *Future Internet*, vol. 12, no. 8, p. 126, Jul. 2020.

KAROL BLIŽŇÁK received the degree in management from Comenius University in Bratislava, where he is currently pursuing the Ph.D. degree with the Faculty of Management. He is an experienced AI and an Analytics Practitioner. His research interests include data governance and its applications in artificial intelligence, business process intelligence, and big data analytics.

MICHAL MUNK received the M.S. degree in mathematics and informatics and the Ph.D. degree in mathematics from Constantine the Philosopher University in Nitra, Slovakia, in 2003 and 2007, respectively. In 2018, he was a Professor of system engineering and informatics with the Faculty of Informatics and Management, University of Hradec Králové, Czech Republic. He is currently a Professor with the Department of Informatics, Constantine the Philosopher University in Nitra. He is also the Head of the Knowledge Discovery Research Group (KDR). His research interests include data analysis, web mining, and natural language processing applications.

ANNA PILKOVÁ has been a Professor of management with the Faculty of Management, Comenius University in Bratislava, since September 2009. Before that, she worked in top managerial positions at a commercial bank in Slovakia, as an Associate Professor with the University of Economics in Bratislava, and chairs the Department of Strategy and Entrepreneurship, Faculty of Management, Comenius University in Bratislava. She runs entrepreneurship, strategic, and risk management courses at bachelor's, master's, and Ph.D. levels. She has been the National Team Leader of the Global Entrepreneurship Monitor, since 2011. She also leads and participates in other international and national projects on entrepreneurship, in which she has conducted research on entrepreneurial activities, entrepreneurship inclusivity (including women, seniors, youth, and migrants), cross-generational entrepreneurship, restart, start-up promotion at HEI institutions, banking regulation, and strategic management in commercial banking in developing countries.