

Trajectories and Challenges for Data Governance Implementation in the Public Sector: A systematic review

Autoria

PATRICIA GOMES RÊGO DE ALMEIDA - patricia.almeida68@yahoo.com

LOYANE MOTA FERNANDES - MOTALOYANE@GMAIL.COM

Programa de Especialização em Orçamento e políticas públicas / UnB - Universidade de Brasília

Marcelo de Oliveira Ribas - marcelo.ribas@gmail.com

Prog de Pós-Grad em Admin – PPGA / UnB - Universidade de Brasília

Daniela Gomes Pessoa - danipessoal@hotmail.com

Prog de Pós-Grad em Admin – PPGA / UnB - Universidade de Brasília

Programa de Pós-Graduação em Administração - PPGADM / UFG - Universidade Federal de Goiás

Mario Gustavo Aquino - mariog.aquino@outlook.com

Propaga / UnB - Universidade de Brasília

Carlos Denner - carlosdenner@gmail.com

57 / UnB - Universidade de Brasília

Resumo

Data governance is a critical aspect of digital transformation and plays a significant role in the implementation of digital government in public administration. The systematic review of the literature presented in this article aims to investigate which research themes are shown in the literature on "data governance in the public sector." The analysis identified the presence of several key streams/themes of literature centered around: the initiatives for collaboration between institutions, the dissemination of information to citizens, the implementation of smart cities, the "datalization" of people and the protection of their data; the development and use of artificial intelligence following ethical principles, and the diagnosis of the current situation of digital transformation in the public sector. Having analyzed all these themes of research, this paper organizes the challenges faced when implementing data governance in public organizations and proposes an integrative agenda to reinforce current and create new streams of investigation that could identify the skills and competencies necessary to involve all actors in an agile, ethical public organization for data governance, as well as, the investigation of necessary peculiar characteristics of each data governance approach in each key theme cited above.

Trajectories and Challenges for Data Governance Implementation in the Public Sector: A systematic review

ABSTRACT

Data governance is a critical aspect of digital transformation and plays a significant role in the implementation of digital government in public administration. The systematic review of the literature presented in this article aims to investigate which research themes are shown in the literature on "data governance in the public sector." The analysis identified the presence of several key streams/themes of literature centered around: the initiatives for collaboration between institutions, the dissemination of information to citizens, the implementation of smart cities, the "datalization" of people and the protection of their data; the development and use of artificial intelligence following ethical principles, and the diagnosis of the current situation of digital transformation in the public sector.

Having analyzed all these themes of research, this paper organizes the challenges faced when implementing data governance in public organizations and proposes an integrative agenda to reinforce current and create new streams of investigation that could identify the skills and competencies necessary to involve all actors in an agile, ethical public organization for data governance, as well as, the investigation of necessary peculiar characteristics of each data governance approach in each key theme cited above.

Keywords:

Data governance; Big Data; Data quality; Digital transformation; Digital government

INTRODUCTION

Data governance has been cited by digital transformation and digital governance strategies worldwide. Despite being pointed out as a success factor in such initiatives in government agencies, Data Governance has not been seen with a high level of maturity in these organizations (Erkut, 2020), requiring many debates on how to implement it. The topic often follows the discussions of data scientists whose protagonist has grown exponentially in recent years due to the growing search for data-based decisions.

Because digital transformation has been presented as a holistic change process with the need for the engagement of the entire organization (Hausberg, 2019), and being dependent on data governance, this article seeks to investigate what approaches have been given by academia for the theme "data governance in the public sector." Would such approaches be restricted to the universe of information technology and data science?

THEORETICAL REFERENCES

Like any broad process of change, the digital transformation of public institutions requires risk management across numerous variables that transcend technology. The new digitized processes depend on correct, complete and accurate data involving different actors. Data governance becomes a fundamental part of organizations' success by admitting data as crucial for digital transformation (Omar, 2020).

Recognized as a set of formal processes in which roles and responsibilities are established in actions related to corporate data (Haneem, 2019), throughout its cycle - creation, storage, use, evolution, archiving and deletion (Omar, 2020), Data Governance aims to maximize the value of data in the decision-making processes that are based on it, minimize operational and decision-

making risks, and protect data from the misuse (Khatri, 2016; Abraham, 2018). This requires designing and implementing data quality strategies to support the organization's business needs (Carretero, 2017; Benfeldt, 2020). In this context, data quality is recognized when the data perfectly serves the purposes of its consumers (Benfeldt, 2020).

During the data lifecycle, challenges can be identified concerning the data itself (e.g. truthfulness, quality, variety), about processes (data capture techniques, data integration, how to choose model analysis appropriate data) and about management, such as privacy, security, governance and ethical aspects (Sivarajah, 2017).

Analyzing itself as a macro process, data governance has as inputs: legislation, national and international regulations, organizational strategy, digital strategy, geographic and departmental distribution of decisions, level of harmonization of administrative processes, IT architecture and organizational culture. After its implementation, it produces the following results: roles and responsibilities, clarity of place of decision-making, data strategy, policies, standards, processes, contractual agreements, indicators, compliance monitoring and problem management (Abraham, 2018). The procedural nature of data governance requires the awareness and training of all actors involved in digital processes. This task has been a significant challenge in public institutions, sometimes only prioritized through compliance mechanisms (Omar, 2020; Khatri, 2016).

Given its scope across the entire organization, data governance is a necessarily collaborative initiative in which each actor does their part to benefit a more expansive universe. This understanding leads to some predictable behaviours in actions requiring collaboration, such as some classic cases from game theory, where there are conflicts between prioritizing comfort and contributing to the benefit of a group (Benfeldt, 2020).

One of the research focuses on data governance occurs when trying to identify, model, or evolve organizational or interorganizational processes; a situation in which distinct administrative units must be considered, not restricted to Information Technology processes (Thomas, 2019), which implies a broad approach throughout the organization, understanding how people act, or should act, concerning data. Consequently, the changes involve the abandonment of management and action based on silos of knowledge areas and the adoption of a collaborative effort between the IT unit and the business units (Vilminko-Heikkinen, 2020). As the benefits obtained by organizations moving towards digital transformation are proportional to the quality of the data they use, ensuring data quality requires planning, strategy and resources for the company (Thomas, 2019), making no sense in an isolated process to maintain data quality (Carretero, 2017).

Given the enormous volume and types of data, prioritizing the treatment of master data - representatives of the organization's primary business elements - becomes a facilitator for implementing data governance. Master data management includes creating and maintaining this data to avoid duplication and increase quality and integration (Haneem, 2019).

Motivated by the digitization of their processes, public organizations have invested in the search for maturity in data management. One of the most adopted assessment models is the Data Management Maturity Model (DMM, 2014), segmented into processes for data management strategy, data governance procedures, data quality methods, techniques for data operations, and processes for architecture and platforms (Thomas, 2019).

In multiple case studies of big data in public organizations, Niño (2020) identified data governance as the lowest maturity factor compared to infrastructure, data management and data structure. Admitting data as strategic assets (Abraham, 2018), low maturity in data governance, carried out in public institutions in any sphere of government, increases the risk of its performance

in the digital sphere. The likely impacts of such risks can range from inefficiency to the complete loss of credibility of the public service provided or failure of national security in the context of digital government.

One of the consequences of the digital transformation in associations is the reflection of facts that occurred in the physical dimension being observed in the data sphere. As this perception evolves, there is an advance in the selected data models and the governance and management process of the implemented data (Khatri, 2106), which materializes the evolution of maturity in the subject.

METHODS

Although the theme of data governance was introduced in the literature after 2007 (Carretero, 2017), concerns about its need gained comprehensive media coverage in general, when researchers and influential people began to brag about the risks of using artificial intelligence when reliable data are not utilized (Musk, 2014; Bostrom, 2014). According to Almeida et al. (2021), the increase in the number of publications related to the regulation of artificial intelligence resulting from its misuse was only identified from 2015 onwards. Considering the combination of the facts described, the period from 2016 to 2021 was chosen for this study.

To guarantee the integrity and validity of the collected information, this research carried out a systematic literature review, based on the Preferred Reporting Items approach for Systematic Reviews and Meta-Analysis - PRISMA (Page, 2021). The eligibility criteria for selecting the articles in this research were divided into two stages:

- 1) Type of publication: Scientific articles published in journals classified by the Journal Citation Reports (JCR) as quartile Q1 and quartile Q2 were selected to ensure a higher quality of publications to be considered in the systematic review.

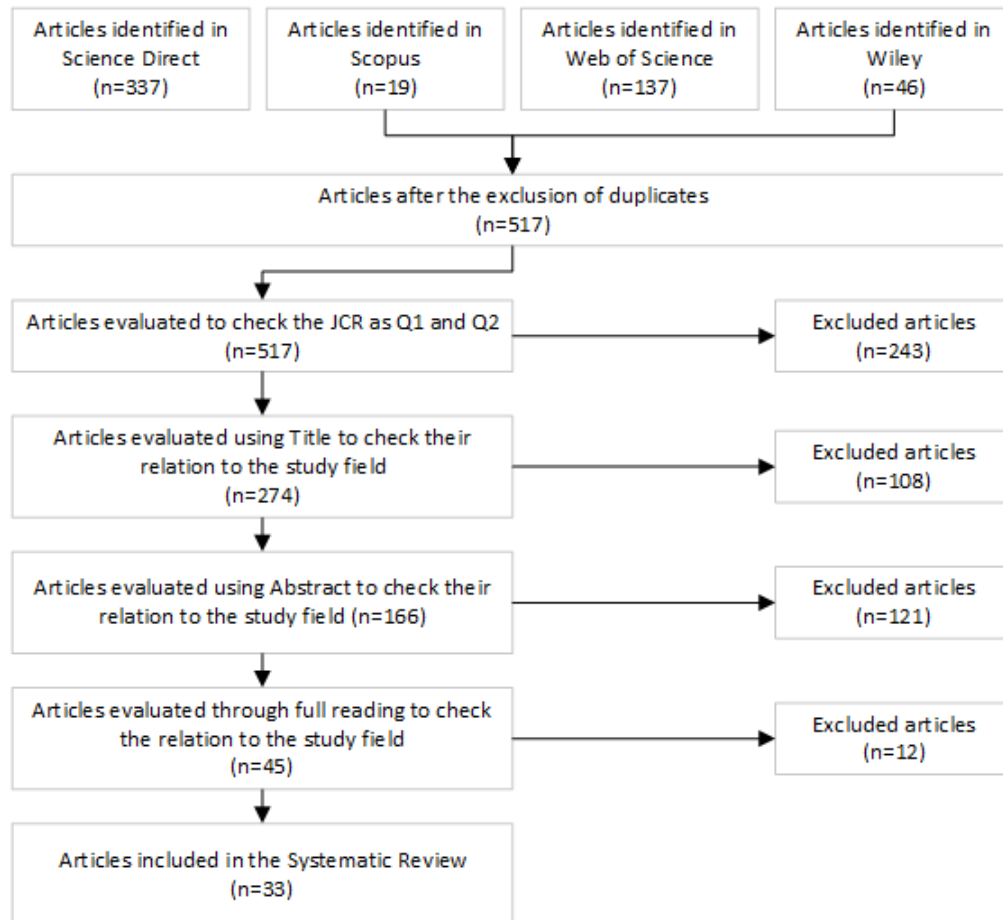
2) Area of study: Scientific articles related to Data Governance. In this study, "Data Governance" is considered as a subset of information governance, which can be defined as formal processes in which roles and responsibilities are established on actions related to corporate data (Haneem, 2019), throughout its cycle - creation, storage, use, evolution, archiving and deletion (Omar, 2020). It seeks to maximize the value of data in the decision-making processes based on it and protect data from misuse (Khatri, 2106). This requires the design and implementation of data quality strategies to support the organization's business needs (Carretero, 2017).

In the literature identification stage, the following terms were defined as descriptors: ("DATA GOVERNANCE") AND ("GOVERNMENT" OR "PUBLIC"), which were researched in the four main research bases in the field of applied social sciences: Web of Science (WOS), Science Direct, Wiley and Scopus.

In these search bases, the search terms were searched in all available fields (title, abstract, keyword) and considered only scientific articles published in English, from January 2016 to September 2021, related to sub-area "business management," "computer science," and "sociology."

In the screening phase, articles published in high-impact scientific journals classified by the Journal Citation Reports (JCR) between quartiles Q1 and Q2 were selected for the first time. Then, articles that dealt with a subject other than data governance excluded from the records based on the "title," "abstract," and reading of the full text were identified. Finally, all records were aggregated, resulting in 33 articles included in the systematic review. The article selection process is detailed in Figure 1.

FIGURE 1.
PRISMA flowchart for the literature review process.



RESULTS

The selection procedure

Two types of approaches/research themes were observed in the sample about the motivating factors of the "data governance" theme. In the first type, the researchers, already aware of the importance of the theme, used the study to deepen representation models, identify maturity assessment criteria, identify factors that generate conflicts in their support and even strategies for implementing data governance in public organizations.

The second part of the researchers delved into researching other themes and identified data governance as a critical path to success in different situations:

- a) initiatives for collaboration between institutions and dissemination of information to citizens;
- b) implementation of smart cities, “datalization” and personal data protection;
- c) development and use of artificial intelligence following ethical principles.

Collaborative initiatives between institutions and dissemination of information to citizens

The need to join data-oriented collaborative actions has intensified (Ruijter, 2021). It allows for richer analyses, comparative studies, completes inter-institutional digital flows and generates data in a volume necessary for artificial intelligence.

The reach of such actions can extend from the transversal sharing between sectors (public and private) (Ruijter, 2021; Vassilakopoulou, 2018) between public institutions, extending even between nations. In this latter case, data domains are created through which solid national strategies can be established. The United States, the European Union and China are clear examples of these domains, with data-driven economies. Rules are selected based on data for the international market, national security and cyber security (Aaronson, 2018).

In these data ecosystems, benefits are based on results that would not be achieved if each party acted separately, even with financial resources (Vassilakopoulou, 2018). However, the care taken to ensure that the data is reliable and uninterrupted is indisputable (van den Broek, 2018). A breach of trust in these inputs would lead to a loss of confidence in the partnerships, in addition to the expectation of a loss of quality and credibility of the services that depend on them.

Data governance is presented in several surveys as necessary to sustain these collaborative actions, given a proper name as decision and management processes and structures that involve people in data-based initiatives beyond the borders of public agencies, levels of government, public or private spheres for a purpose that would not be achieved without such an agreement (Ruijter,

2021). Some researchers have studied the reasons for the need for quality assurance (Vassilakopoulou, 2018), data protection and integrity; from the risks of its inexistence (Aaronson, 2018), to the benefits that public institutions and nations start to have when they implement governance processes of these shared data (van den Broek, 2018).

Other researchers go further by exposing the expanding challenges for creating and sustaining data governance processes when faced with data resulting from collaboration between different public institutions and even nations. These cases assume a much greater level of complexity than governance in a single public organization, in which problems of low quality, integrity and protection pose risks of breaking the financial system, the national development chain and even the security and sovereignty of one or more countries (Aaronson, 2018). Complexity is also reflected in the search for long-term sustainable collaborations on a global scale (Vassilakopoulou, 2018).

Data-driven collaborative initiatives can feature centralized and distributed data governance (Micheli, 2020). The second case results in greater operational complexity firmly based on trust (Micheli, 2020), while the first imposes strict agreements between the parties involved, as occurs in European Union agreements (Aaronson, 2018).

An example of complex discussions involves collecting, curating, quality, and access control of genetic data, whose research involves hospitals and laboratories. This scenario imposes transparency of processes and individual results, requiring insertion in rigid data governance models associated with the medical regulation of the countries involved. This equation consists of transparency (opening the bases) and people's privacy, copyrights and patents (Vassilakopoulou, 2018).

Concerning data opening, the extreme side occurs when this opening is given to any internet user, common in digital governments that motivate innovation based on data reuse (Ruijter, 2021; van den Broek, 2018). Opening bases to the world's citizens is one of the most robust arguments for accountability in public organizations and nations, as well as enablers of world development (Zhao, 2021).

Implementation of smart cities, “datalization” of people and data privacy

A specific type of data sharing takes place in public services offered in the concept of smart cities. Smart cities are increasingly frequent in digital government proposals as data-based solutions to social problems (Artyushina, 2020; Calzada, 2020), characterized by many integrated digital services to improve citizens' quality of life. Its citizens, forming an ecosystem of city data (Gupta, 2020). In this way, transport, traffic control, health (prevention, treatment and emergency care), energy supply, public safety and education are the most prioritized services in the digitization proposals (Kolesnichenko, 2021).

When talking about smart cities, the most controversial issue is the combination of two factors: 1) the data that support them are from people who live in or transit through these urban centers. They are, therefore, personal data or resulting from the behaviour of people; 2) to handle large volumes of data and an extensive portfolio of digitally advanced services, the governments of these cities often establish contracts and partnerships with large private technology companies (Artyushina, 2020).

Based on this combination, it is observed that people's data started to be stored and processed by private companies, members of digital production chains in different regions of the world. Complaints about the sale of data to other private sector companies without notification or authorization from the people to whom they refer are not uncommon (Artyushina, 2020).

Given this scenario, the need for a new data governance model has been demanded by researchers, and the concept of data ownership is once again discussed, as well as rules, processes and even regulations that guarantee the privacy and security of residents and passers-by in these cities (Artyushina, 2020; Popham, 2018). A balance is sought between what can be automated and what needs to be avoided. One of the strategies suggested for the smart cities government to have more strength to impose contractual rules is the association of cities with standard practices in their public policies served by digital public services (Calzada, 2020) or the promotion of laboratory cooperation agreements of innovation with the government (Gupta, 2020).

Considering the growing advancement of digital services, infiltrating from the most basic habits and needs of people (Calzada, 2020) to their exercise as a citizen (Barassi, 2020), it is often not necessary to be inserted in a city designed and adapted under the smart city label to be immersed in a context of "data dating" of people (Calzada, 2020; Popham, 2018), where the data governance processes of public services become imperative for the lives of these individuals in social, psychological and financial aspects (Barassi, 2020; Rhahla, 2021).

In cases of bolder digitization and excessive centralization of information in non-democratic governments, the combination of smart city initiatives and "datalization" people alert them to significant risks to human rights (Calzada, 2020).

The situations described reinforces the need to make the population and public managers aware of the risks and advantages of the massive use of data so that, when possible, decisions are taken in a conscious manner (Kolesnichenko, 2021)

Development and use of artificial intelligence in line with ethical principles

The growing use of digital services and interconnected devices (internet of things) has been identified as drivers of large data repositories in the most diverse formats (Andrews, 2018).

Organized independently or collaboratively, these large volumes of data are fertile ground for using artificial intelligence (AI) (Calzada, 2020). Equally growing has been the rhetoric of public managers regarding AI expectations in government agencies (Dwivedi, 2021; Medaglia, 2021), associated with greater efficiency, productivity (Dwivedi, 2021), interaction with the citizen (Medaglia, 2021), but also related to people's quality of life (Alshahrani, 2021) and the global sustainability of the environment (Dwivedi, 2021).

Often presented in the context of interpreting and learning data skills to achieve specific goals (Dwivedi, 2021), systems using AI can be defined as those that exhibit intelligent behaviour through analysis of the environment and actions, with some level of autonomy, to achieve specific goals (AI HLEG, 2019). In a more practical approach, AI can also be understood as technologies that combine algorithms, data and processing power to replicate the human cognitive process - learning, reasoning and autocorrelation (Alshahrani, 2021).

Given the reliance of artificial intelligence on data, a strong relationship between data quality and integrity and the outcome of AI-based digital services. This is because the predictive models that underlie AI algorithms - a set of rules to be followed when calculating or solving a problem (Mergel et al., 2016) - are more accurate in terms of the volume of data to be trained and the level the reliability of these data increases (Janssen, 2020).

However, there are some variables involved in the lifecycle of an AI-based service that is potentially susceptible to compromising the trust and security of its users. These are risks arising from the data used by the model, causing unwanted effects on the development, commercialization and inappropriate use of AI. They can be perceived as prejudice, discrimination (Vetrò, 2021; Alshahrani, 2021), loss of privacy, false positive and false negative, loss of autonomy, harm (psychological, financial or physical), reduced human rights, unfair judgments, the concentration

of power (Kuziemski, 2020), not excluding others. There has often been the need to defend against fake news through a reliable data ecosystem that allows for the correct explanation of the fake fact (Andrews, 2018).

To mitigate such risks, AI-based systems need to be reliable, fair, transparent (Janssen, 2020) and their results explainable by the data (Dwivedi, 2021) (Andrews, 2018). The motivation for the explanatory character of the data for autonomous decision-making imposes accountability actions on public institutions that develop systems based on AI or provide services that use AI (Janssen, 2020), despite the recognition of the difficulty in giving them this transparency, when you want to comply with legal requirements for the protection of personal data (Kuziemski, 2020).

In addition to technological issues, risks can also be mitigated through a data governance model that includes policies, regulations (Medaglia, 2021) and that directs the institution towards culture and actions of equality, justice, transparency (Vetro, 2021) (Medaglia, 2021), freedom of information and ethically accepted principles, as well as the protection of confidential and personal data (Janssen, 2020). As a result, risk assessment processes and ethical principles (Vetro, 2021) become relevant instruments (Janssen, 2020) in Data Governance (Kuziemski, 2020) (Alshahrani, 2021; Medaglia, 2021), sometimes intersecting with AI Governance (Medaglia, 2021; Andrews, 2019), presented by Dwivedi (2021) as processes that guarantee autonomous systems the correct and expected value of their actions.

In orientation to prepare for any legal problems arising from failures or unwanted results in the use of their AI-based services, and mitigation of the risks mentioned above, public organizations were instructed, some already with legal requirements, to structure their data ecosystem (Kuziemski, 2020), to formalize and standardize its data governance processes

(Medaglia, 2021; Kuziemski, 2020), with clear rules on the responsibilities of all actors necessary for the management and use of data throughout its data cycle. life (Janssen, 2020).

Bibliometric analysis

The grouping of selected articles by year of publication shows an increase in the number of publications on Data Governance in the public sector. The 2021 value was not higher than 2020 because the survey took place until September 2021. In addition, publications carried out in 2021 were possibly negatively affected by the covid-19 pandemic due to logistics difficulties between those involved until authors, reviewers, and editors have adjusted to the new production processes.

FIGURE 2
Number of publications with propositions or hypothesis validation per year

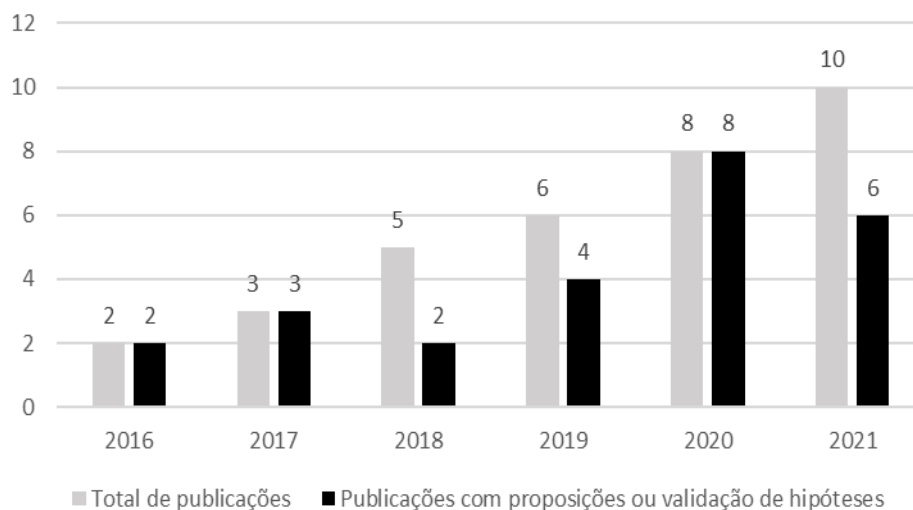
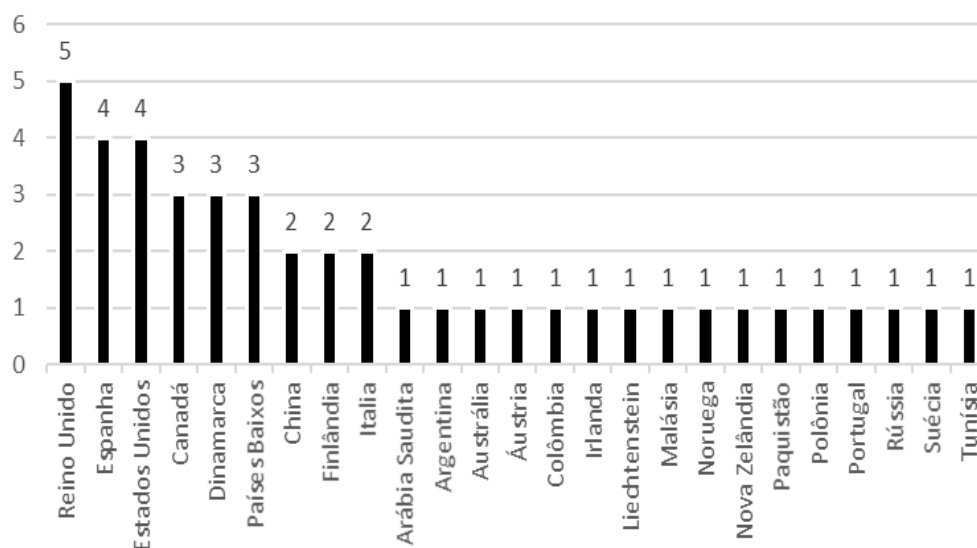


FIGURE 3
Distribution of the number of articles published based on author's country



An analysis was also carried out regarding the structure of the selected articles, taking into account whether they were restricted to discussing the importance of the existence of data governance or whether, in addition, the proposed models for better understanding and evolution of governance or sought to validate hypotheses on the subject. With this last type of approach, it is observed in figure 2, in general, the growth of these publications over the years, except for only 2021, as the collection took place until September.

As shown in the figure 3, for the countries of origin of the researchers the existence of one or more authors from each country in each publication was computed as an occurrence. So, for example, a publication with two authors from the United Kingdom was calculated as an occurrence. In contrast, another publication with an author from Spain and one from Canada led to calculating an occurrence for each of these countries. Thus, several countries ($n = 44$) are more significant than the number of articles analyzed ($n = 33$). Although the data shows that the United Kingdom has more authors publishing, the data shows considerable geographic diversity as to the origin of authors, with sixteen countries with a selected author. This diversity indicates that the

concern with the topic is not concentrated in one or a few countries but geographically diverse, with some predominance of European countries.

Regarding the areas of knowledge of the selected articles, there was, as expected, a predominance of authors from the areas of Administration and Information Technology, with 30% of the works exclusively for each of these areas. 21% of the works with authors from Administration and Information Technology were also gathered. The remaining 18% also had authors from other areas of knowledge: Law, Social Sciences and Communication, and studies that brought together Information Technology researchers with other areas of expertise (Mathematics, Medicine, Administration, Human Sciences, Law).

TABLE 1
Articles included in the systematic review

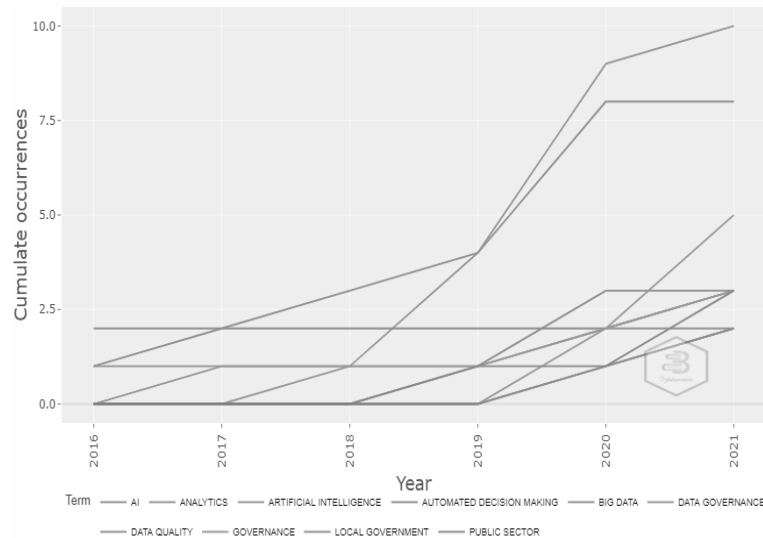
<i>Autor's study locus</i>		<i>Approach type</i>		<i>Research Methods</i>	
Management	30%	Qualitative	88%	Case study	17
IT	30%	Quantitative	6%	Bibliographic research	14
IT and management	21%	Mixed methods	6%	Documentary research	8
Others	18%			Field Research	5
				Quantitative methods	4
				Survey	3
				Multiple-case	2
				Ethnography research	2

As for the approach used, of the 33 selected studies, 88% used qualitative research methods, 6% quantitative methods and another 6% mixed methods.

The analysis techniques used were quite varied within the possible reach of qualitative research. Case studies, multiple cases, documentary research, field research, ethnographic research and the survey stand out, allowing a critical look at the different approaches that public organizations have given to data governance.

In a more in-depth analysis, using the free software Bibliometrix R-package from the Biblioshiny application, the data represented in Graph 1 were found.

GRAPH 1
Thematic Keyword Evolution Map from 2016 to 2021



GRAPH 2
Thematic Keyword Evolution Map from 2016 to 2021



The first graph (Graph 1) shows the evolution of the themes found, based on the keywords of the selected articles in the last five years. In general, a positive trend (ascending or of stability) was observed in all identified themes. In the analysis of the most frequent keyword among the articles, it was noted that the expression "data governance" had an expected growth from 2018 on

an ascending curve to 2021, which indicates a growing emergence of the theme for the coming years. In turn, the word "big data" was the second most frequent keyword, appearing in publications from 2017 onwards and reaching its stability point between 2020 and 2021, with no growth trend. It was also observed that the keyword "artificial intelligence" gained prominence in publications only from 2019 onwards and, since then, with a significant growth trend.

Extending the analysis of keywords to the way they relate to the entire sample, it is observed that clusters, in turn, are also linked to other clusters. In Graph 2, only four clusters closest to each other were selected for display. For this analysis, the different groups that are far from the ecosystem "data governance," "big data," and "artificial intelligence," and "smart cities" will be disregarded. It is possible to observe that among the themes with the highest occurrences, that is, those with the largest bubble size in each point of the network, there are only the themes "data governance" and "big data." The four closest keyword groupings are presented in Table 2.

TABLE 2
The keywords description

<i>Keyword</i>	<i>Description</i>
<i>Data governance</i>	Presents a cluster of very broad terms: "data privacy", "surveillance capitalism", "coerced digital participation", "data literacy", "self-efficacy", "design similarity model", "cognitive social theory", "Digital transformation", "digital renunciation", in addition to some locus expressions
<i>Big data</i>	In its ecosystem, contains technological themes, mainly: "quality measurement model in use", "data policy", "digital platform", "data infrastructure", "leadership public", "public administration", in addition to the terms associated with the locus
<i>Artificial intelligence</i>	Although less frequent, it has strong links with "big data" and "data governance". Some linked terms in your ecosystem - also follow these links: "data management", "data quality", "maturity model", "automated decision making" and "data ethics"

Smart cities Features ties to "data literacy", "community engagement" and "digital invisibility"

DISCUSSION

The growing number of publications and the countries where the authors' research centers are distributed on four continents allows us to consider the perception of the theme's importance as global. Equally broad is the amount of knowledge areas that also aroused for such a study, breaking the bias that Data Governance is an exclusive subject of Information Technology (30% of the sample).

The proportions prove the various reports in the sample in which the procedural, behavioural and regulatory nature of Data Governance is emphasized and the technological dimension. The high proportion of the sample (51%) coming from administration (with and without IT) may represent the different forms of corporate or state collaboration initiatives; opening the institutions' information to citizens; case studies of the conditions for implementing smart cities, and concerns about regulation; and application of ethical principles in the production, marketing and use of artificial intelligence. The surveys in multiple areas (18%) may reflect the concern with the protection of personal data (Law), sociological aspects and the improvement of the quality of life of people with smart cities.

The excellent quality of the articles selected in the sample, according to the selection criteria of Q1 and Q2 journals (Fig. 1), may have been the reason for the high and growing proportion of studies that presented model proposals to understand better, address the challenges or improve the data governance, not limited to just arguing its need (fig 2a).

It is noteworthy the high proportion of studies with a qualitative approach (88%), which can demonstrate several phenomena: a) the combination of facts considered in this study to define the period of the sample – from 2016 on; b) the long learning curve in all the lines of action required for data governance; and c) the procedural and transversal nature of data governance, which imposes a long negotiation and implementation time. Combined, these phenomena signal a still exploratory scenario, therefore, with little empirical quantitative research, which is in line with the evolution of the themes addressed over time (fig 1a), in which the expression "data governance" only acquires a significant rise, the from 2018, and still maintains a positive inclination, being only accompanied by "big data," whose behaviour has already stabilized.

The evolution in which the themes have been addressed by the studies of the sample (fig 1a) also reveals that the choice of the research period (2016-2021) is coherent with the maturity in which data governance is investigated in public organizations. If maintained over the next few years, the significant trend in the frequency of the expression "artificial intelligence" will meet with "data governance."

Automated keyword grouping reveals how closely the themes are related and the study's strategy concerning data governance. In the group around "data governance," the closest links highlight social, managerial, technological and behavioural aspects that support the motivations for implementing data governance. This range demonstrates alignment to the sample subgroups described at the Results Section as well as the items: Collaborative initiatives between institutions and dissemination of information to citizens; implementation of smart cities, "datalization" of people and data privacy and development and use of artificial intelligence in line with ethical principles . But what stands out the most is the number of links with the different groups, confirming the different approaches of the sample.

In the grouping adjacent to "big data," the number and arrangement of its connections demonstrate how it fits specialists who want to go deeper into models or strategies for better implementation and optimization of data governance.

The grouping around "artificial intelligence," in turn, although less frequent, has strong links with "big data" and with "data governance." Some terms linked in their ecosystem also follow these links, present semantics associated with ethics and technology, and are very close to the studies identified at subsection "Development and use of artificial intelligence in line with ethical principles".

The grouping around "smart cities," in turn, aligns with the studies identified at "Implementation of smart cities, "datalization" of people and data privacy".

It is essential to highlight how the automatic analysis of keywords carried out in the texts of the articles resulted in groupings very close to those identified by the careful reading of each one of them. Therefore, by the two analysis strategies, it can be confirmed that the term data governance was, in fact, the main focus of the research, assuming relationships with the others mentioned. Furthermore, it is closer to groupings aligned with the themes identified in subsections: "Collaborative initiatives between institutions and dissemination of information to citizens; implementation of smart cities, "datalization" of people and data privacy and development and use of artificial intelligence in line with ethical principles" and more technical or technical-managerial themes, such as digital platforms and digital transformation.

Because of the analyzes that resulted in the themes registered in the subsections above mentioned, it is, therefore, possible to recognize the need to implement data governance in situations where they exist: initiatives for collaboration between institutions and for opening up information to citizens; implementation of smart cities, "datalization" of people and protection of

personal data; and development and use of artificial intelligence, taking into account ethical principles, in addition to the general situation of digital transformation in public institutions.

One can observe a common purpose in all three approaches cited: a huge concern about a responsible use of data.

From this perception, it is understood that the study and involvement in data governance become imperative not only for those whose object of research and investment is data or information but also academic and practical actors in the following contexts:

TABLE 3
Characteristics of data governance context

<i>Approach</i>	<i>Description</i>
<i>Initiatives for collaboration between institutions and dissemination of information to citizens</i>	Planning, designing, financing and managing collaboration agreements, partnerships, contracts involving services or digital information in which at least one public organization, government, consortium among nations is engaged
	Management of public organizations, governments in any sphere, parliaments and courts of law, with the production or use of digital services using personal data
	Planning, design, financing, and management of public organizations in which information is wanted or needed to be made available to citizens
<i>Implementation of</i>	Planning, designing, funding and implementing digital transformation strategies in public organizations or digital government evolution
	Planning, design, financing, management, partnerships, commercialization of digital services aimed at smart cities

<i>smart cities, “datalization” and personal data protection</i>	<p>Planning, design, financing, and management of a company or state agency, which has a part in the creation, provision or consumption of services in smart cities</p> <p>Public bodies involved with national security and sovereignty</p> <p>Public bodies investigating fraud and any crimes whose object is on a digital model</p>
<i>Development and use of artificial intelligence following ethical principles</i>	<p>Planning, designing, financing, producing, marketing and using services that contain artificial intelligence in public organizations</p> <p>Public bodies investigating fraud and any crimes whose object is on a digital model</p>

The situations described in Table 3, fitting into scenarios of immersion in the digital world. However, not all are classified as such, except for the last case. They are not always presented with the necessary depth to the actors so that risk minimization and value enhancement strategies can occur satisfactorily.

Contributions and Implications for Research: an agenda

This study allowed us to clarify the perception that situations not directly involved with the study of data science or data governance require the implementation of well-established data governance processes. The main beneficiaries of this study are the academic community and professionals engaged in: collaborative initiatives between institutions and dissemination of information to citizens, ; smart cities projects; organizations’ capabilities and processes through personal data analytics; development and use of artificial intelligence under ethical principles, in addition to the general situation of digital transformation in public institutions.

The research has limitations regarding the eligibility criteria for selecting the articles since the exclusion of articles published in journals not classified as Q1 and Q2 removes from the analysis 243 articles that could indicate other research perspectives in Data Governance. In addition, a possible limitation is the analysis of thematic trends based on keywords and not on the content of selected papers.

The long time required to implement Data Governance and the large number and diversity of actors involved were the most cited arguments in the studied sample. Considering the permanence of these statements in articles published throughout the sample period, an agenda is proposed to develop proposals for the use of models that use the Agile philosophy for this purpose, as well as studies that can identify technical and competence skills and competencies—non-technical, necessary for all actors involved for different types of public organizations.

Focusing on amplifying the benefits of data governance implementations, further researches can investigate peculiar characteristics of each of the three situations described at the Results session, supporting a deep and specialized approach.

CONCLUSIONS

This study aimed to investigate how Data Governance in public organizations is being approached by academia. We sought to contemplate multiple bases and additional areas of study for information technology through a systematized review. The results showed approximately 70% of the sample published by research centers linked to not IT areas. In addition to the classic study approach in identifying problems, new models, or ways of evolving data governance, the following approaches were identified in the studies: initiatives for collaboration between institutions and for opening up information to citizens; implementation of smart cities, "datalization" of people and

protection of personal data; development and use of artificial intelligence in compliance with ethical principles, in addition to the general situation of digital transformation in public institutions.

The research objective was achieved by identifying situations specific to action in the public sphere (alone or in partnership/contract with the private sector), which, despite not being immersed in the data area, require study and deepening in the implementation of data governance.

If we associate the responsible use of data to the decisions made around the world, this study highlights that actors of different fields of knowledge need to concern how the data are being used (or will be used) to rule our lives.

REFERENCES

- Aaronson, S. A., & Leblond, P. 2018. Another Digital Divide: The Rise of Data Realms and its Implications for the WTO. *Journal of International Economic Law*, 21(2), 245–272. <https://doi.org/10.1093/jiel/jgy019>
- Abraham, R., Schneider, J., & vom Brocke, J. 2019. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49(July), 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Alshahrani, A., Dennehy, D., & Mäntymäki, M. 2021. An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Government Information Quarterly*, (February). <https://doi.org/10.1016/j.giq.2021.101617>
- Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. 2021. Artificial Intelligence Regulation: A framework for governance. *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-021-09593-z>
- AI HLEG - High-Level Expert Group on Artificial Intelligence. 2019a. para a Comissão Europeia. *A Definition of AI: Main Capabilities and Disciplines. Definition developed for the AI HLEG's deliverables*.
- Amann, J., Sleight, J., & Vayena, E. 2021. Digital contact-Tracing during the Covid-19 pandemic: An analysis of newspaper coverage in Germany, Austria, and Switzerland. *PLoS ONE*, 16(2 February), 1–16. <https://doi.org/10.1371/journal.pone.0246524>
- Andrews, L. 2019. Public administration, public leadership, and public value construction in the age of algorithm and big data. *Public Administration*, 97(2), 296–310. <https://doi.org/10.1111/padm.12534>
- Artyushina, A. 2020. Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics*, 55(July), 101456. <https://doi.org/10.1016/j.tele.2020.101456>
- Austin, Lisa M. e Lie, D. 2019. Safe Sharing Sites. *NYUL Rev. Forthcoming*, 94.
- Barassi, V. 2019. Datafied Citizens in the Age of Coerced Digital Participation. *Sociological*

- Research Online*, 24(3), 414–429. <https://doi.org/10.1177/1360780419857734>
- Benfeldt, O., Persson, J. S., & Madsen, S. 2020. Data Governance as a Collective Action Problem. *Information Systems Frontiers*, 22(2), 299–313. <https://doi.org/10.1007/s10796-019-09923-z>
- Borgogno, O., & Colangelo, G. 2019. Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law and Security Review*, 35(5), 105314. <https://doi.org/10.1016/j.clsr.2019.03.008>
- Bostrom, N. 2014. *Superintelligence: Paths, strategies and dangers*. Oxford: OUP.
- Calzada, I., & Almirall, E. 2020. Data ecosystems for protecting European citizens' digital rights. *Transforming Government: People, Process and Policy*, 14(2), 133–147. <https://doi.org/10.1108/TG-03-2020-0047>
- Clark, N. E., & Chongtay, R. 2020. Technological Mediation for Disaster Risk Management. *Journal of Contingencies and Crisis Management*, 28(4), 411–415. <https://doi.org/10.1111/1468-5973.12331>
- Carretero, Ana; Gualo, Fernanda; Caballero, Ismael; Piattini, M. 2017. MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. *Elsevier BV*, 54, 139–151. Retrieved from <https://doi.org/10.1016/j.csi.2016.11.008>
- Combata Niño, H. A., Combata Niño, J. P., & Morales Ortega, R. 2020. Business intelligence governance framework in a university: Universidad de la costa case study. *International Journal of Information Management*, 50(December 2017), 405–412. <https://doi.org/10.1016/j.ijinfomgt.2018.11.012>
- Creswell, J. W. 2007. *Projeto de pesquisa: método qualitativo, quantitativo e misto*. Porto Alegre: Artmed.
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Williams, M. D. 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57(August 2019), 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Erkut, B. 2020. From Digital Government to Digital Governance: Are We There Yet? *Sustainability*, 12(3), 860. <https://doi.org/10.3390/su12030860>
- Elizabeth A. Evans, Elizabeth Delorme, K. C. and D. M. G. 2020. A qualitative study of big data and the opioid epidemic: recommendations for data governance. *BMC Medical Ethics*, 21.
- Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., Yeong, Z. K. 2021. Governing AI safety through independent audits. *Nature Machine Intelligence*, 3(7), 566–571. <https://doi.org/https://doi.org/10.1038/s42256-021-00370-7>
- Gleeson, N., & Walden, I. 2016. Placing the state in the cloud: Issues of data governance and public procurement. *Computer Law and Security Review*, 32(5), 683–695. <https://doi.org/10.1016/j.clsr.2016.07.004>
- Graef, I., & Prüfer, J. 2021. Governance of data sharing: A law & economics proposal. *Research Policy*, 50(9), 104330. <https://doi.org/10.1016/j.respol.2021.104330>
- Gupta, A., Panagiotopoulos, P., & Bowen, F. 2020. An orchestration approach to smart city data ecosystems. *Technological Forecasting and Social Change*, 153(January), 119929. <https://doi.org/10.1016/j.techfore.2020.119929>
- Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Abu Bakar, N. A. 2019. Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management*, 45(April 2018), 25–43.

- <https://doi.org/10.1016/j.ijinfomgt.2018.10.007>
- Hannah Rebutischa, Caroline Thompsonb, Laurence Côté-RoycS, arah M. 2020. Unicorn planning: Lessons from the rise and fall of an American ‘smart’ mega-development. *Cities*, 101. <https://doi.org/10.1016/j.cities.2020.102686>
- Hausberg, J. P., Liere-Netheler, K., Packmohr, S., Pakura, S., & Vogelsang, K. 2019. Research streams on digital transformation from a holistic business perspective: A systematic literature review and citation network analysis. *Journal of Business Economics*, 89(8), 931–963. <https://doi.org/10.1007/s11573-019-00956-z>
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. 2020. Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 37(3), 101493. <https://doi.org/10.1016/j.giq.2020.101493>
- Jiang, G., Cai, X., Feng, X., & Liu, W. 2021. Effect of data environment and cognitive ability on participants’ attitude towards data governance. *Journal of Information Science*. <https://doi.org/10.1177/016555152111019000>
- Kastouni, M. Z., & Ait Lahcen, A. 2020. Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.11.024>
- Khatri, V. 2016. Managerial work in the realm of the digital universe: The role of the data triad. *Business Horizons*, 59(6), 673–688. <https://doi.org/10.1016/j.bushor.2016.06.001>
- Kolesnichenko, O., Mazelis, L., Sotnik, A., Yakovleva, D., Amelkin, S., Grigorevsky, I., & Kolesnichenko, Y. 2021. Sociological modelling of smart city with the implementation of UN sustainable development goals. *Sustainability Science*, 16(2), 581–599. <https://doi.org/10.1007/s11625-020-00889-5>
- Kuziemski, M., & Misuraca, G. 2020. AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
- Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. 2021. Artificial Intelligence in Government: Taking Stock and Moving Forward. *Social Science Computer Review*, 1–18. <https://doi.org/10.1177/08944393211034087>
- Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. 2016. A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, 63, 123–130. <https://doi.org/10.1016/j.future.2015.11.024>
- Mergel, I., Rethemeyer, R. K., & Isett, K. 2016. Big Data in Public Affairs. *Public Administration Review*, 76(6), 928–937. <https://doi.org/10.1111/puar.12625>
- Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. 2020. Emerging models of data governance in the age of datafication. *Big Data and Society*, 7(2). <https://doi.org/10.1177/2053951720948087>
- Musk, E. 2014. I Hope Artificial Intelligence Is Nice to Us. *New Perspectives Quarterly*, 31, 51–55. <https://doi.org/10.1111/npqu.11427>
- Nisar, Q. A., Nasir, N., Jamshed, S., Naz, S., Ali, M., & Ali, S. 2020. Big data management and environmental performance: role of big data decision-making capabilities and decision-making quality. *Journal of Enterprise Information Management*, 34(4), 1061–1096. <https://doi.org/10.1108/JEIM-04-2020-0137>
- Omar, A., & Almaghthawi, A. 2020. Towards an integrated model of data governance and integration for the implementation of digital transformation processes in the Saudi Universities. *International Journal of Advanced Computer Science and Applications*,

- 11(8), 588–593. <https://doi.org/10.14569/IJACSA.2020.0110873>
- Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71): 1-6. doi:10.1136/BMJ.n71
- Popham, J., Lavoie, J., & Coomber, N. 2020. Constructing a Public Narrative of Regulations for Big Data and Analytics: Results From a Community-Driven Discussion. *Social Science Computer Review*, 38(1), 75–90. <https://doi.org/10.1177/0894439318788619>
- Rhahla, M., Allegue, S., & Abdellatif, T. 2021. Guidelines for GDPR compliance in Big Data systems. *Journal of Information Security and Applications*, 61, 102896. <https://doi.org/10.1016/j.jisa.2021.102896>
- Ruijter, E. 2021. Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly*, 38(4), 101612. <https://doi.org/10.1016/j.giq.2021.101612>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Tamburri, D. A. 2020. Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and evaluation. *Information Systems*, 91, 101469. <https://doi.org/10.1016/j.is.2019.101469>
- Thomas, M. A., Cipolla, J., Lambert, B., & Carter, L. 2019. Data management maturity assessment of public sector agencies. *Government Information Quarterly*, 36(4), 101401. <https://doi.org/10.1016/j.giq.2019.101401>
- van den Broek, T., & van Veenstra, A. F. 2018. Governance of big data collaborations: Balancing regulatory compliance and disruptive innovation. *Technological Forecasting and Social Change*, 129(November 2016), 330–338. <https://doi.org/10.1016/j.techfore.2017.09.040>
- Vassilakopoulou, P., Skorve, E., & Aanestad, M. 2019. Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Information Systems Journal*, 29(4), 768–786. <https://doi.org/10.1111/isj.12191>
- Vetrò, A., Torchiano, M., & Mecati, M. 2021. A data quality approach to identifying discrimination risk in automated decision-making systems. *Government Information Quarterly*, (February), 101619. <https://doi.org/10.1016/j.giq.2021.101619>
- Vilminko-Heikkinen, R., & Pekkola, S. 2019. Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, 47(April 2018), 76–87. <https://doi.org/10.1016/j.ijinfomgt.2018.12.017>
- Wang, Y., Kung, L. A., Gupta, S., & Ozdemir, S. 2019. Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organizations: A Configurational Perspective. *British Journal of Management*, 30(2), 362–388. <https://doi.org/10.1111/1467-8551.12332>
- Zhang, R., Indulska, M., Sadiq, S., Zhang, J., Li, S., Wang, Y., Andrews, L. 2020. Public administration, public leadership, and public value construction in the age of the algorithm and 'big data.' *BMC Medical Ethics*, 14(1), 1–17. <https://doi.org/10.3389/fpubh.2021.655447>
- Zhao, Y., & Fan, B. 2021. Understanding the key factors and configurational paths of the open government data performance: Fuzzy-set qualitative comparative analysis. *Government Information Quarterly*, 38(3), 101580. <https://doi.org/10.1016/j.giq.2021.101580>