International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-9, Issue-9; Sep, 2022
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.99.21

Data-Driven Decision Making in the Public Sector: A Systematic Review

Rodrigo Speckhahn Soares da Silva, Claudelino Martins Dias Junior, Rogério Tadeu de Oliveira Lacerda

The Graduate Program in Administration, Federal University of Santa Catarina, Brazil

Received: 11 Aug 2022,

Received in revised form: 07 Sep 2022,

Accepted: 12 Sep 2022,

Available online: 17 Sep 2022

***Keywords*— data driven, decision making, public sector, open government.**

***Abstract*— This study presents the selection process of a bibliographic portfolio and bibliometric analysis related to Data-Driven Decision-making (DDD) in the context of Public Administration using the ProKnow-C framework. We proceeded to search and select articles to compose a portfolio and then analyze their characteristics. The journals where the theme is recurrent were evaluated, as well as the authors and articles that stand out. With the dissemination of ICTs and the Open Government movement, it has been noticed that the volume of articles on the subject has grown significantly in the last decade. In this context, the methodology used proves to be a useful tool for building knowledge in a given field of research, providing a structured and rigorous procedure that minimizes the use of randomness and subjectivity in the literature review process.*

I. INTRODUCTION

With the rapid technological changes in data storage and processing, managers and administrators have been changing the way they make decisions. They have relied less on their intuitions and more on data. This change has become necessary, as Jetzek *et al.* (2014) point out, because of the myriad possibilities for creating, collecting, and storing data in our increasingly digital world. According to data from Statista (2022) through 2021, and estimates from 2022 through 2025, the growth in the creation, capture, and consumption of information and data is evident, as presented in Fig. 1.

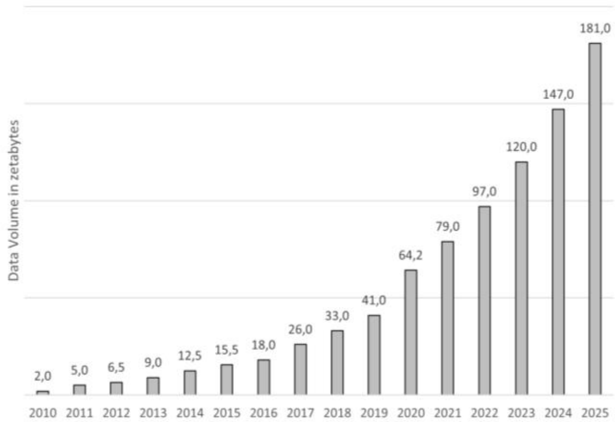

Fig. 1: Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025 adapted from (Statista, 2022)

In this sense, Brynjolfsson and McElheran (2016) conducted a systematic empirical study regarding the diffusion of what has been termed Data-Driven Decision-Making (DDD).

At the industrial level, DDD has been primarily concentrated on the following characteristics: (i) large-scale companies, (ii) owning and using information technology, (iii) having skilled workers, and (iv) significant levels of awareness.

In the scope of Public Administration, typically the owner of large volumes of data and information, information technologies are increasingly used and, despite the different realities and specificities of each state or nation, there are trained professionals in greater or lesser numbers. In addition, there is a movement toward making government data available, known as Open Government Data (OGD).

According to Matheus *et al*. (2020) efforts in this direction can result in more democracy, greater administrative efficiency, transparency, accountability, collaboration, engagement, and trust in government. In addition, it can also potentially result in the generation of considerable economic and social value. However, according to Jetzek *et al.* (2014), there is still a lack of understanding of how this can happen indicating the need for greater attention and further exploration of the topic.

Birchall (2015) states that OGD is part of a necessary component of the new "data economy." To participate and gain benefits from the so-called "*infocapitalism*" democracy, where the data subject is called to be: (i) auditor who monitors granular state transactions in the name of accountability, (ii) entrepreneur who makes data profitable through apps and visualizations, and (iii) consumer of these apps and visualizations.

In this growing context of data and information comes Data Science which is intrinsically intertwined with other concepts of growing importance such as big data, artificial intelligence, and the DDD. This perspective provides a framework and principles that allow the manager to systematically address problems to extract useful knowledge from data and thus make more assertive decisions (Provost & Fawcett, 2013).

Data scientists in the government context need not only a solid knowledge of statistics and data analysis, the use of techniques and tools for predictive purposes and for visualizing results. But also an understanding of other elements such as policy-making, organization, legislation, and public values. This combined knowledge allows the data to be placed in context and to understand its use and the implications involved (Matheus et al., 2020).

II. LITERATURE REVIEW

DDD, as a new paradigm, emerges from digitalization and networks and is based on a new and valuable resource, data. Thus, new practices have been spreading rapidly among companies regardless of organization sizes, as well as in Public Administration (Klingenberg et al., 2019).

Provost e Fawcett (2013) define DDD as the practice of basing decisions on data analysis rather than purely on intuition. They also emphasize that it is not an all-or-nothing practice, meaning that it can be used to a greater or lesser degree within organizations.

In this context, there is also the movement called "Open Government" or "Open Data", defined by Kassen (2013a), as one in which government data is available for use and distribution by anyone without any copyright restrictions. Thus, the dichotomous world (Market and State) has been transformed into an open and interconnected world in which the traditional roles and relationships between these agents are being replaced by complex interdependencies. Therefore, the production and use of these data for decision-making and their availability by public authorities become even more significant to the extent that citizens and public and private organizations have, not only the opportunity but also the motivation and ability to use data to achieve social and economic value (Jetzek et al., 2014).

Brynjolfsson *et al*. (2011) statistically showed that the more a company is data-driven, the more productive it is, and it can achieve gains of around 4% to 6%. They also highlight the correlation (almost causal) with a higher return on assets, equity, asset utilization, and market value. Similarly, they point out that productivity increases in the context of Public Administration when DDD is used, with gains of around 5% to 6% beyond what can be explained by traditional inputs and the use of IT.

Data Science has supported and increasingly overlapped with DDD through automated computational systems. Whether it is decisions for which discoveries need to be made within the data or decisions that are repeated especially on a large scale. Another critical aspect is the support of analytical thinking from data, the reason is that this skill is important for both data scientists and employees across the organization. For it allows one to understand the fundamental concepts and have frameworks to organize analytical thinking. It not only enables interaction with competence but also in visualizing opportunities to improve DDD or to see data-driven competitive threats. However, investments in analytics can be useless, and even harmful, unless employees can also incorporate that data into complex decision-making. Therefore, for Data Science to flourish as a field, it must think beyond the commonly used algorithms, techniques, and tools. It needs to think about the elementary principles and concepts that underlie the techniques and the systematic thinking that promotes success in DDD. The success desired in the DDD business environment requires the ability to think about how the fundamental concepts apply to specific problems and businesses (analytical thinking) (Provost & Fawcett, 2013).

Public value is another related concept in the OGD and e-government literature. The public value framework is based on the premise that public resources should be used to increase value, not only in the economic sense but also more broadly in terms of what is valued by citizens.

2.1. The methodological framework of this study

This is an exploratory study as regards its purpose and bibliographical as regards the means of investigation because it is a systematic study developed based on published articles (Vergara, 2004).

According to Broadus (1987) bibliometrics is a type of quantitative and statistical bibliographic research that originated in Information Science. However, this study is also qualitative because the data obtained will be analyzed according to interests, delimitations, and criteria defined by the authors.

Thus, to obtain a set of bibliographic references about data-driven decision-making in the context of public administration; and from this portfolio which is the articles, authors, and prominent journals dealing with this theme, the structured procedure called Knowledge Development Process - Constructivist (*ProKnow-C*) was used.

The *ProKnow-C* framework starts by considering the researcher's interest in a theme, as well as some delimitations and restrictions that help him, in a structured way, to select and analyze relevant articles. According to Ensslin et al (2010) the concept of bibliometric analysis is based on the quantitative evidencing of the parameters of a selected set of articles: the selected articles themselves, their sets of bibliographic references, authors, relevant journals, and the number of citations..

The next section presents the methodological procedures used in the search, collection, selection, and analysis of relevant publications related to the theme under study.

III. METHODOLOGICAL PROCEDURES

Kitchenham (2004) summarizes that systematic reviews are means to assess and interpret relevant research available for a research question, thematic area, or phenomenon of interest. Among the main motivations for studies of this nature, he highlights the possibility of (i) synthesizing evidence concerning treatment or technology to summarize, for example, empirical evidence of the benefits and limitations of a specific method; (ii) identifying gaps in current research in order to suggest areas for further investigation; (iii) providing a framework to adequately position new research activities and; (iv) assessing the extent to which empirical evidence may support or contradict theoretical hypotheses, or assist in the generation of new hypotheses.

Karlsson (2009), regarding the use of systematic reviews, highlights (i) the scientific support when basing work on relevant publications; (ii) justify the choice of a theme and the consequent contribution of a research proposal; (iii) substantiate the methodological framework of the research; (iv) by delimiting the scope of research, the researcher, makes it feasible and; (v) allows the researcher to develop his analytical capacity of the information and criticism of the specific literature.

3.1. The filters

Thus, as to the procedures adopted in this study, the procedures described in the sequence were carried out in the months of May and June 2022.

Two databases were selected as sample fields. The base *Web of Science* (or ISI) gives rise to the JCR index (*Journal Citation Report*) that evaluates the impact factor of journals and the base Scopus (Elsevier) which currently holds the title of the largest database of scientific articles in the world.

The first filter for the selection of articles was the choice of keywords grouped into two thematic axes: "data-driven" and "public sector". The keywords and their respective synonyms initially selected relative to each axis are presented in Table 1.

Table.1: Selected keyword combinations

Nº	*Axis 1*	*Axis 2*
1	"data-driven*"	"Public*"
2		"Public Admin*"
3		"Public Sec*"
4		"Public Serv*"
5		"Public Manage*"
6		"Govern*"
7		"Open Gov*"
8		"Open Data*"

Synonyms and wildcard characters were used so as not to restrict too much the search results in the databases. The searches with these words were carried out only in the titles, keywords defined by the authors of the articles, and in the abstract of the articles ("TOPIC" selection in the search fields of the databases).

Only articles (type: "ARTICLE") published in the last 10 years were selected, that is, published from 2012 until June 2022, when this research was carried out.

The areas of interest selected in each base are presented in Table 2.

Table.2: Selected areas of interest on each basis

Web of Science	***Scopus***
“Mathematics Interdisciplinary Applications”, “Management”, “Business”, “Business Finance”, “Economics”, “Interdisciplinary Applications”, “Public Administration”, “Management”, “Multidisciplinary Sciences”, “Social Sciences Mathematical Methods” e “Social Sciences Interdisciplinary”	“Business”, “Management and Accounting”; “Economics”, “Econometrics and Finance”, “Decision Sciences”; “Social Sciences” e “Multidisciplinary”.

These were the preliminary filters adopted in the search for each keyword combination in each database. The next section will present the results and respective analyses conducted.

IV. PRESENTATION AND DISCUSSION OF RESULTS

The search results for each keyword combination, on each basis, are shown in Table 3.

Table.3: Database search results

Combinations	***Web of Science***	***Scopus***
“data-driven*” AND "Public*"	264	888
“data-driven*” AND "Public Admin*"	21	18
“data-driven*” AND "Public Sec*"	19	41
“data-driven*” AND "Public Serv*"	12	41
“data-driven*” AND "Public Manage*"	9	11
“data-driven*” AND "Govern*"	600	615
“data-driven*” AND "Open Gov*"	9	20
“data-driven*” AND "Open Data*"	31	83
Total	**965**	**1.717**

From the total of articles obtained in each base, 745 duplicate articles were identified, resulting in a total of 1.937 distinct articles.

The next step was to read the titles of the articles in search of articles aligned with the theme of interest. After this step, 1.633 articles were excluded, i.e., 304 were aligned with the proposed theme.

We then proceeded to analyze the scientific recognition of these 304 articles. For this, using the Google Scholar (*GS*) tool, the number of citations of each article was obtained.

As a cut-off criterion, the articles that represent around 80% of the total number of citations (8.599) were selected. Thus, of the 304 articles aligned by title 56 (or 18.43% of the total) concentrated 80,044% of the citations. In other words, the articles that received at least 38 citations were selected.

The 248 less cited articles will still be evaluated according to other criteria, and some may still be part of the final portfolio of articles selected as part of the theoretical framework of the research.

With the articles with the greatest scientific recognition, they were evaluated as to the alignment of the abstract with the theme of interest. In this process, 11 non-aligned articles were eliminated.

Thus, 45 articles remained that were aligned as to the title and the abstract, which presented a relevant quantity of citations

The 248 articles with few or no citations were also evaluated according to the following criteria: (i) articles published less than 2 years before the analysis, since there was not enough time to be cited yet; and (ii) when published more than 2 years before, being from researchers who are already among the authors of the 45 articles selected so far.

Among the 248 articles under review, 174 were published in 2020, 2021, or 2022. And among the 74 articles with a publication date of more than 2 years, only 1 was by an author present in the bibliographic portfolio.

After reading the abstracts of these 175 articles, 6 were selected based on their alignment with the research objective.

Thus, these 6 articles were added to the set of 45 previously selected for further reading. After the full reading of the 51 articles, 7 were excluded for being misaligned with the research theme, resulting in a set of 44 articles for the final portfolio.

In summary, the results obtained in the first two stages of the framework (Search and Selection) are shown in Fig. 1. And the results of the bibliometric analysis itself (Analysis) will be presented in the next section.

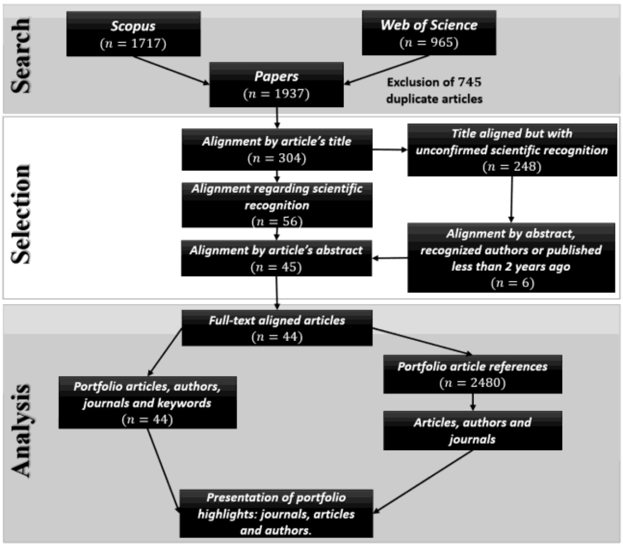

Fig. 1: Main steps of ProKnow-C framework adapted from Lacerda et al. (2016)

The 44 articles in the portfolio are presented in alphabetical order of the first author in Table 4.

Table.4: Articles from the bibliographic portfolio

Nº	*Article*	*Nº of citations in GS*
1.	(Provost & Fawcett, 2013)	1.477
2.	(Williamson, 2016)	413
3.	(Kassen, 2013)	362
4.	(Chakraborty & Ghosh, 2020)	293
5.	(Jetzek et al., 2014)	248
6.	(Liang et al., 2018)	218
7.	(Bansak et al., 2018)	210
8.	(Barns, 2018)	205
9.	(Elish & Boyd, 2018)	196
10.	(Matheus & Janssen, 2020)	149
11.	(Parycek et al., 2014)	137
12.	(Phillips-Wren & Hoskisson, 2015)	121
13.	(Chen et al., 2017)	90
14.	(Appelbaum et al., 2018)	88
15.	(Khalifa et al., 2014)	81
16.	(Birchall, 2015)	76
17.	(Batarseh & Latif, 2016)	73
18.	(Tenney & Sieber, 2016)	72
19.	(Klingenberg et al., 2019)	69
20.	(McBride et al., 2019)	66
21.	(Katsonis & Botros, 2015)	61
22.	(Kassen, 2017)	59
23.	(Choi et al., 2018)	58
24.	(Hino et al., 2018)	53
25.	(Gupta & Rani, 2019)	51
26.	(Moro Visconti & Morea, 2019)	49
27.	(Poel et al., 2018)	48
28.	(Waheed et al., 2018)	48
29.	(Marda, 2018)	47
30.	(Matheus et al., 2020)	43
31.	(van Oort et al., 2015)	43
32.	(Toufaily et al., 2021)	42
33.	(Kassen, 2018)	41
34.	(Lourenço et al., 2017)	41
35.	(French, 2014)	39
36.	(Dencik et al., 2019)	38
37.	(Hummel et al., 2021)	38
38.	(Severo et al., 2016)	38
39.	(M. Janssen et al., 2022)	18
40.	(Pereira et al., 2018)	14
41.	(Kassen, 2020)	6
42.	(Kim et al., 2019)	6
43.	(Chen & Ji, 2022)	0
44.	(Cheung & Chen, 2021)	0

4.1. Bibliometric analysis of the bibliographic portfolio

This section is dedicated to the bibliometric analysis of the selected portfolio to build a theoretical framework for data-driven decision-making in the context of Public Administration. The results will be presented in three stages: (i) a bibliometric analysis of the articles selected; (ii) a bibliometric analysis of the references of the articles in the portfolio; and (iii) the classification of the articles according to their relevance to the scientific community.

From the bibliometric analysis, the journals with the largest number of articles are shown in Fig. 2.

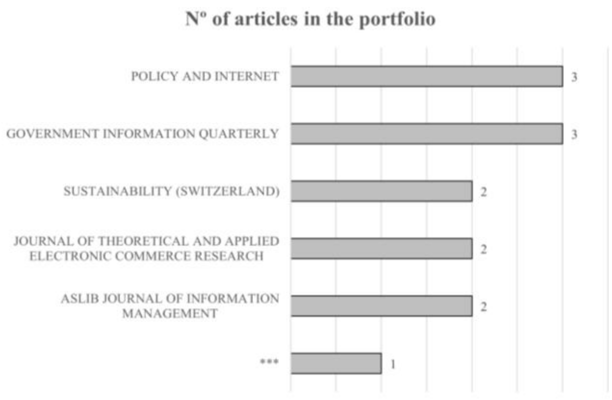

Fig. 2: Journals with the highest number of articles in the portfolio

The journals "*Policy and Internet*" and "*Government Information Quarterly*" presented 3 articles, each one, among those selected for the portfolio. The other periodicals (*Annals of Operations Research, Australian Journal of Public Administration, Behaviour and Information Technology, Big Data, Big Data and Society, Big Data Research Chaos Solitons & Fractals City Culture and Society, Communication Monographs, European Journal of Social Theory, Information & Management, Information Technology and People, International Journal of Disaster Risk Reduction, International Journal of Electronic Government Research, Internet Policy Review, Journal of Accounting Literature, Journal of Decision Systems, Journal of Education Policy, Journal of Information Science, Journal of Manufacturing Technology Management, Law and Social Inquiry, Nature Sustainability, Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, Public Performance & Management Review, Public Transport, Science, Social Science Computer Review, Surveillance and Society, Transforming Government: People Process and Policy, Transportation Research Part D: Transport and Environment, Urban Education* e *Urban Planning*), which appear, indicated with *** in Fig. 2, contributed only 1 article each.

The authors who stood out within the portfolio with the highest number of articles are shown in Fig. 3.

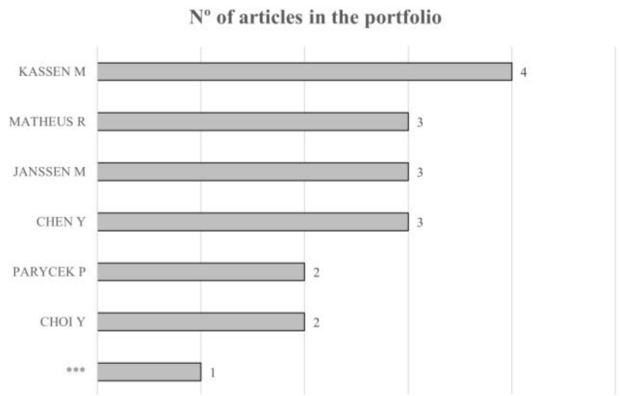

Fig. 3: Authors with the highest number of articles in their portfolio

Researcher Maxat Kassen from Nazarbayev University (Kazakhstan) had 4 papers selected for the final portfolio, followed by Ricardo Matheus, Marijn Janssen, and Chen Yang with 3 papers, and Peter Parycek and Youngseok Choi with 2 papers each. The remaining 101 authors had only one of their papers selected.

As for scientific recognition, by the number of citations in GS, the articles are presented in descending order in Fig. 4.

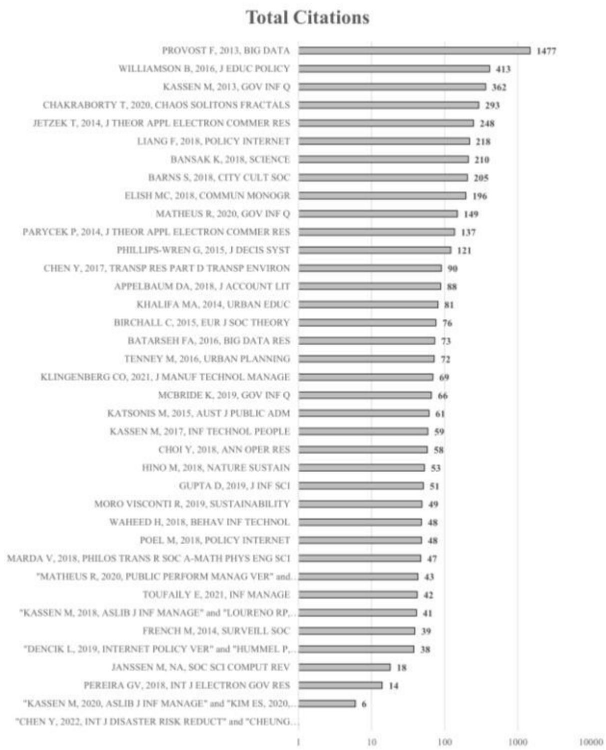

Fig. 4: Portfolio articles are ordered by the number of citations

The most prominent article, with 1,477 citations, is the one by Foster Provost and Tom Fawcett entitled "*Data*

Science and its Relationship to Big Data and Data-Driven Decision Making".

In time, within the final portfolio, the occurrences of keywords indicated by the authors were analyzed. The results are presented in Fig. 5.

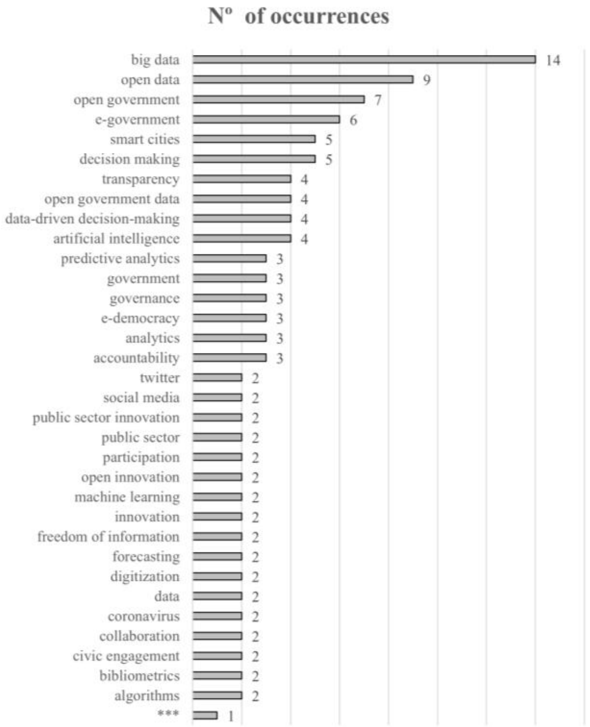

Fig. 5: Keywords most frequently used by authors in the articles in the portfolio

We identified 33 keywords cited at least twice by the authors. The most frequent was "*big data*" with 14 mentions, followed by "*open data*" with 9 mentions and "*open government*" with 7, and "*smart cities*" and "*decision-making*" with 5 mentions each. In addition to these, another 156 keywords were mentioned only once by the authors of the portfolio, as can be seen qualitatively in Fig 6.

Fig. 6: Wordcloud with the author's keywords

4.2. Bibliometric analysis of references from the bibliographic portfolio

From the 44 articles in the portfolio, 2.480 different references were obtained. As for the analysis of the most frequent journals in the portfolio references, the results are shown in Fig. 7.

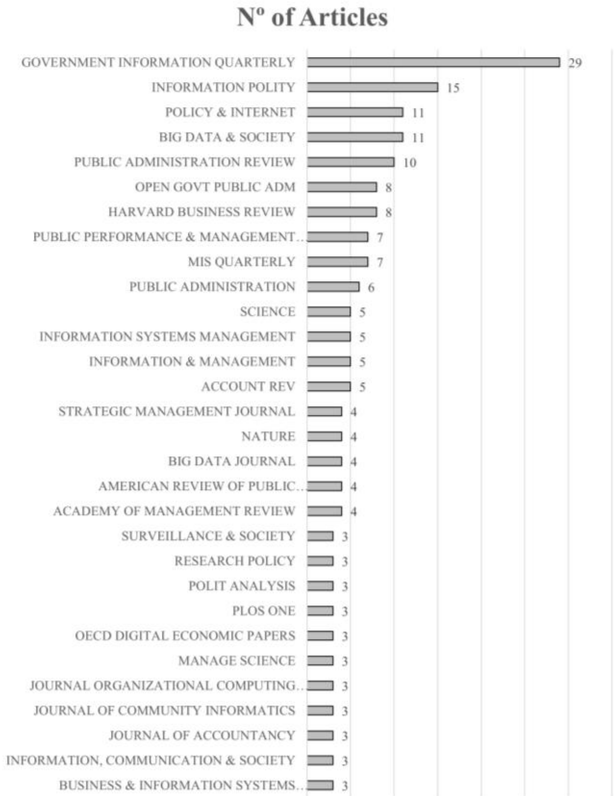

Fig. 7: Most Relevant Periodicals in the Portfolio References

These are the 30 most prominent journals among the references found in the portfolio. The most prominent journal is "*Government Information Quarterly*" with 29 occurrences, followed by "*Information Polity*" with 15 occurrences and "*Policy & Internet*" and "*Big Data*" with 11 occurrences each.

Among the 3.374 authors cited by the articles in the portfolio, 34 authors stood out with 6 or more citations. These authors are shown in Figure 8.

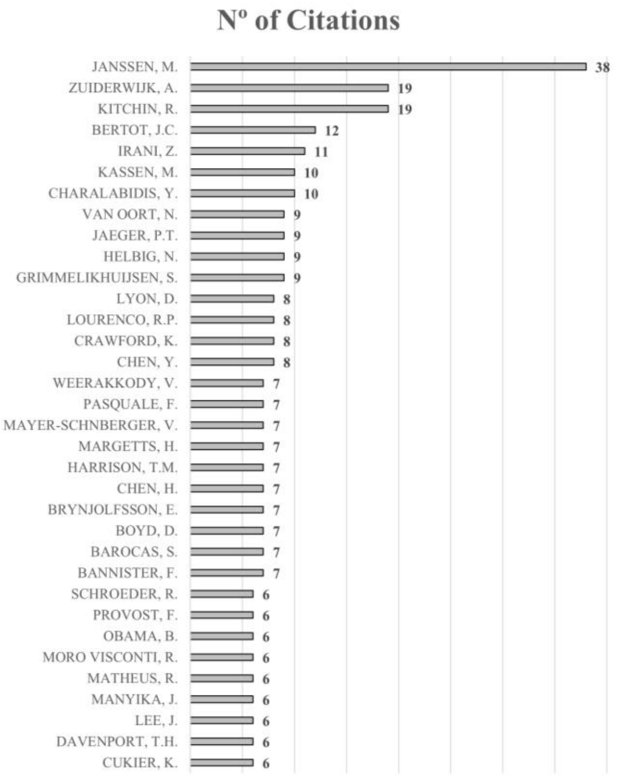

Fig. 8: Most cited authors among the references of the articles in the bibliographic portfolio

The most prominent author in the portfolio references is Marijn Janssen from the Delft University of Technology with 38 citations, followed by Anneke Zuiderwijk and Rob Kitchin with 19 citations each, and John Bertot with 12 citations.

The 20 most prominent articles (number of citations in the GS in the portfolio references are shown in Table. 5.

Table.5: Most prominent articles among those cited in the portfolio references

Nº	*Article*	*Nº of citations*
1.	(Boyd & Crawford, 2012)	5.189
2.	(Albino et al., 2015)	3.255
3.	(Bertot et al., 2010)	2.846
4.	(Kitchin, 2014b)	2.675
5.	(Kitchin, 2014a)	2.666
6.	(Kitchin, 2014c)	2.279
7.	(M. Janssen et al., 2012)	2.053
8.	(Burrell, 2016)	1.514
9.	(Butler, 2013)	645
10.	(Gurstein, 2011)	616
11.	(K. Janssen, 2011)	380
12.	(Kassen, 2013)	378
13.	(Lourenço, 2015)	318
14.	(Gonzalez-Zapata & Heeks, 2015)	259
15.	(M. Janssen & Zuiderwijk, 2014)	211
16.	(Peled, 2011)	200
17.	(Clarke & Margetts, 2014)	186
18.	(M. Janssen & Kuk, 2016)	83
19.	(Barocas, Solon; Selbst, Andrew D, 2016)	60
20.	(Kashin et al., 2015)	7

The article "*Critical Questions for Big Data*" by Danah Boyd and Kate Crawford, both contributors at Microsoft Research, was the most cited in *GS* among the 2,480 articles in the bibliographic references in the portfolio.

Among the 109 authors of the portfolio presented, in Fig. 9, the 39 authors presented at least 2 articles in the portfolio and 1 in the portfolio references.

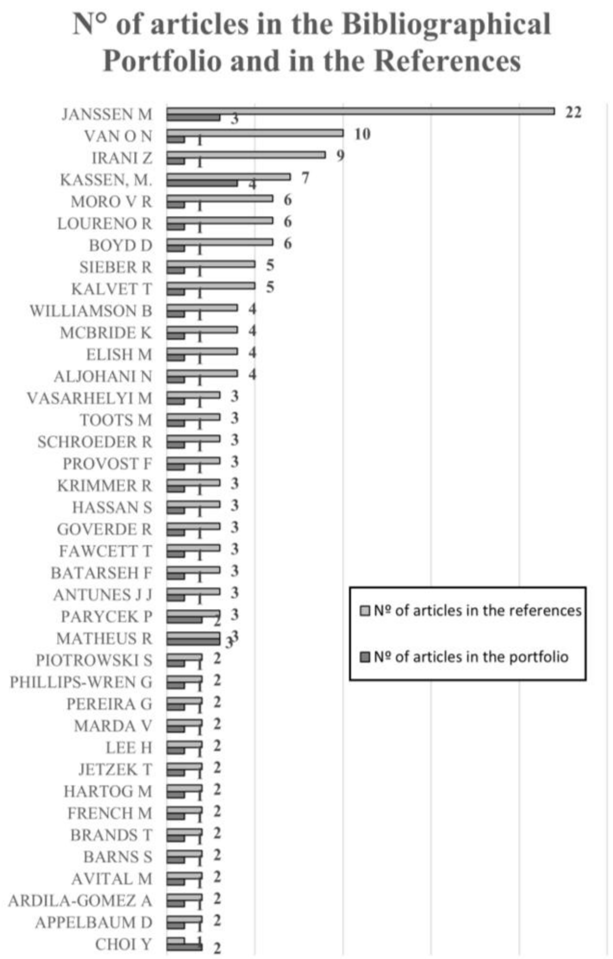

Fig. 9: Authors with the highest number of articles in the bibliographic portfolio and the references in the bibliographic portfolio

Again, the author Marijn Janssen appears as the leading author among the references with 22 articles. Niels van Oort, also from the Delft University of Technology, with 10 articles, and Zahir Irani, from the Business School of Brunel University, with 9 articles. And Maxat Kassen, the most prominent author in the portfolio (with 4 articles, Fig. 3) had 7 articles cited in the references.

From the bibliometric analysis, the most relevant journals and articles in academia can be evidenced through the combined analysis between the journals where the articles in the portfolio were published and the journals in the references, as shown in Fig. 10.

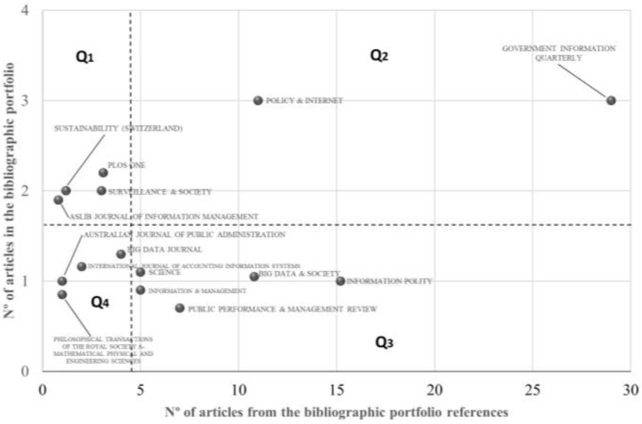

Fig. 10: Top journals in the bibliographic portfolio and references

Fig. 10 was divided into 4 quadrants, in quadrant Q1 we observe the prominent journals in the portfolio ("*ASLIB Journal Information Management*", "*Plos One*", "*Surveillance & Society*" e "*Sustainability*") all with 2 articles each in the bibliographic portfolio. In Q_2 are the periodicals that stand out in the portfolio and in the portfolio references ("*Government Information Quarterly*" and "*Policy & Internet*"), which together have almost 45% of the citations in the portfolio references and 3 articles each in the bibliographic portfolio. In Q_3 the journals that stood out in the portfolio references ("*Big Data & Society*", "*Information & Management*", "*Information Polity*", "*Public Performance & Management Review*" and "*Science*",) with at least 5 citations. And in Q_4 the relevant periodicals in the portfolio and the references of the portfolio ("*Australian Journal of Public Administration*", "*Big Data Journal*", "*International Journal of Accounting Information Systems*" and "*Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*") with one article in the portfolio each and less than 5 citations in the references of the bibliographic portfolio.

When analyzing the scientific relevance of the articles (obtained by the number of citations of each article) and the incidence of articles by the same author in the bibliographic portfolio references, we obtained the scatter plot shown in Fig. 11.

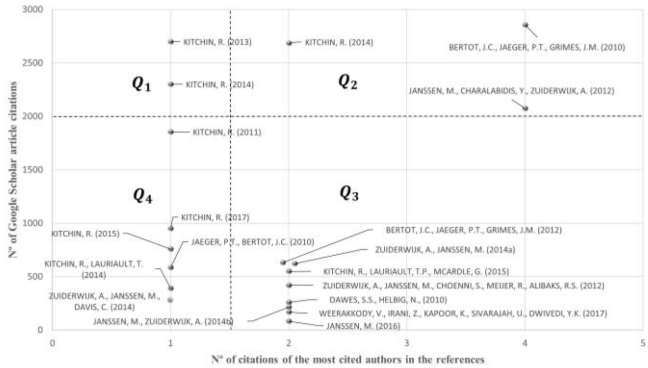

Fig. 11: Top articles and authors in the bibliographic portfolio

Fig. 11 was also divided into 4 quadrants, in quadrant Q_1 are the top articles in terms of citations in *GS* (Kitchin, 2014b) and (Kitchin, 2014c) both with more than 2,000 citations. In Q_2 are the prominent articles performed by prominent authors (Bertot et al., 2010), (M. Janssen et al., 2012) and (Kitchin, 2014a) that got more than 2,000 citations in the *GS* and more than 2 citations in the portfolio. In Q_3 are articles by prominent authors (Bertot et al., 2012), (Dawes & Helbig, 2010), (M. Janssen & Kuk, 2016), (M. Janssen & Zuiderwijk, 2014), (Kitchin et al., 2015), (Weerakkody et al., 2017), (Zuiderwijk et al., 2012) and (Zuiderwijk & Janssen, 2014) all with 2 citations in the references and less than 1000 citations in the GS. And in Q_4 the articles relevant to the topic (Jaeger & Bertot, 2010), (Kitchin & Dodge, 2011), (Kitchin & Lauriault, 2014), (Kitchin, 2017), (Kitchin, 2015) and (Zuiderwijk et al., 2014).

V. CONCLUSION

The objective of this study focuses on the use of a systematized procedure to select relevant articles to compose a theoretical framework about DDD in the context of Public Administration, given the relevance and timeliness of the topic and the economic and social impacts.

This study initially presented the procedures for searching and selecting relevant articles and an analysis to assess the main works, authors, and journals that have been published on the topic. As summarized in Fig. 1, using the *ProKnow-C* framework, from an initial volume of 1,937 articles we obtained a bibliographic portfolio composed of 44 articles presented in Table 4.

In addition to the article selection process, which aims to compose a theoretical referential on the theme, a bibliometric analysis was carried out. It was possible to highlight the journals "*Policy and Internet*" and "*Government Information Quarterly*" as the most prominent in terms of publications on the theme.

As for the authors, the framework evidenced the contributions of Maxat Kassen, Ricardo Matheus, Marijn Janssen, and Chen Yang, with at least 3 papers each.

Furthermore, from the analysis of the bibliographic references of the articles in the portfolio, it was verified the relevance of the journals "*Government Information Quarterly*", "*Information Polity*", "*Policy & Internet*" and "*Big Data*".

Thus, the use of data and new Big Data and Artificial Intelligence technologies and the creation of transparency through OGD initiatives are key areas of research on DDD. This systematic review allowed us to verify the increase in the production of studies related to open data, transparency, and the use of new technologies to treat data, classify, and group data, helping the public manager to obtain insights and make decisions with the help of technical and quantitative elements.

However, we emphasize that the results presented are limited to the sample of journals and articles researched because they cannot be extrapolated to the entire set of publications in an area.

As a suggestion for further and future work, we recommend the application of the next stage of the *ProKnow-C* framework, which proposes a systemic content analysis of this bibliographic portfolio.

REFERENCES

[1] Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology*, *22*(1), 3–21. https://doi.org/10.1080/10630732.2014.942092

[2] Appelbaum, D. A., Kogan, A., & Vasarhelyi, M. A. (2018). Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics. *Journal of Accounting Literature*, *40*(1), 83–101. https://doi.org/10.1016/j.acclit.2018.01.001

[3] Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, *359*(6373), 325–329. https://doi.org/10.1126/science.aao4408

[4] Barns, S. (2018). Smart cities and urban data platforms: Designing interfaces for smart governance. *City, Culture and Society*, *12*, 5–12. https://doi.org/10.1016/j.ccs.2017.09.006

[5] Barocas, Solon; Selbst, Andrew D. (2016). *Big Data's Disparate Impact*. https://doi.org/10.15779/Z38BG31

[6] Batarseh, F. A., & Latif, E. A. (2016). Assessing the Quality of Service Using Big Data Analytics. *Big Data Research*, *4*, 13–24. https://doi.org/10.1016/j.bdr.2015.10.001

[7] Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, *27*(3), 264–271. https://doi.org/10.1016/j.giq.2010.03.001

[8] Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2012). Promoting transparency and accountability through ICTs, social media, and collaborative e-government. *Transforming Government: People, Process and Policy*, *6*(1), 78–91. https://doi.org/10.1108/17506161211214831

[9] Birchall, C. (2015). 'Data.gov-in-a-box': Delimiting transparency. *European Journal of Social Theory*, *18*(2), 185–202. https://doi.org/10.1177/1368431014555259

[10] Boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

[11] Broadus, R. N. (1987). Toward a definition of "bibliometrics". *Scientometrics*, *12*(5–6), 373–379. https://doi.org/10.1007/BF02016680

[12] Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1819486

[13] Brynjolfsson, E., & McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, *106*(5), 133–139. https://doi.org/10.1257/aer.p20161016

[14] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

[15] Butler, D. (2013). When Google got flu wrong. *Nature*, *494*(7436), 155–156. https://doi.org/10.1038/494155a

[16] Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals*, *135*, 109850. https://doi.org/10.1016/j.chaos.2020.109850

[17] Chen, Y., Ardila-Gomez, A., & Frame, G. (2017). Achieving energy savings by intelligent transportation systems investments in the context of smart cities. *Transportation Research Part D: Transport and Environment*, *54*, 381–396. https://doi.org/10.1016/j.trd.2017.06.008

[18] Chen, Y., & Ji, W. (2022). Estimating public demand following disasters through Bayesian-based information integration. *International Journal of Disaster Risk Reduction*, *68*, 102713. https://doi.org/10.1016/j.ijdrr.2021.102713

[19] Cheung, A. S. Y., & Chen, Y. (2021). From Datafication to Data State: Making Sense of China's Social Credit System and Its Implications. *Law & Social Inquiry*, 1–35. https://doi.org/10.1017/lsi.2021.56

[20] Choi, Y., Lee, H., & Irani, Z. (2018). Big data-driven fuzzy cognitive map for prioritising IT service procurement in the

public sector. *Annals of Operations Research*, *270*(1–2), 75–104. https://doi.org/10.1007/s10479-016-2281-6

[21] Clarke, A., & Margetts, H. (2014). Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform: Open, Closed, and Big Data in Public Management Reform. *Policy & Internet*, *6*(4), 393–417. https://doi.org/10.1002/1944-2866.POI377

[22] Dawes, S. S., & Helbig, N. (2010). Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency. Em M. A. Wimmer, J.-L. Chappelet, M. Janssen, & H. J. Scholl (Orgs.), *Electronic Government* (Vol. 6228, p. 50–60). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14799-9_5

[23] Dencik, L., Redden, J., Hintz, A., & Warne, H. (2019). The 'golden view': Data-driven governance in the scoring society. *Internet Policy Review*, *8*(2). https://doi.org/10.14763/2019.2.1413

[24] Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

[25] Ensslin, L., Ensslin, S. R., Lacerda, R. T. de O., & Tasca, J. E. (2010). *ProKnow-C, knowledge development process-constructivist* (INPI Patent).

[26] French, M. (2014). Gaps in the gaze: Informatic practice and the work of public health surveillance. *Surveillance & Society*, *12*(2), 226–242. https://doi.org/10.24908/ss.v12i2.4750

[27] Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001

[28] Gupta, D., & Rani, R. (2019). A study of big data evolution and research challenges. *Journal of Information Science*, *45*(3), 322–340. https://doi.org/10.1177/0165551518789880

[29] Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*. https://doi.org/10.5210/fm.v16i2.3316

[30] Hino, M., Benami, E., & Brooks, N. (2018). Machine learning for environmental monitoring. *Nature Sustainability*, *1*(10), 583–588. https://doi.org/10.1038/s41893-018-0142-9

[31] Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1), 205395172098201. https://doi.org/10.1177/2053951720982012

[32] Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*(4), 371–376. https://doi.org/10.1016/j.giq.2010.05.003

[33] Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, *28*(4), 446–456. https://doi.org/10.1016/j.giq.2011.01.004

[34] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

[35] Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-supported Decision-Making in Government. *Social Science Computer Review*, *40*(2), 478–493. https://doi.org/10.1177/0894439320980118

[36] Janssen, M., & Kuk, G. (2016). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. https://doi.org/10.1080/10919392.2015.1124005

[37] Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, *32*(5), 694–711. https://doi.org/10.1177/0894439314525902

[38] Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-Driven Innovation through Open Government Data. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 15–16. https://doi.org/10.4067/S0718-18762014000200008

[39] Karlsson, C. (Org.). (2009). *Researching operations management*. Routledge.

[40] Kashin, K., King, G., & Soneji, S. (2015). Explaining Systematic Bias and Nontransparency in U.S. Social Security Administration Forecasts. *Political Analysis*, *23*(3), 336–362. https://doi.org/10.1093/pan/mpv011

[41] Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

[42] Kassen, M. (2017). Open data in Kazakhstan: Incentives, implementation and challenges. *Information Technology & People*, *30*(2), 301–323. https://doi.org/10.1108/ITP-10-2015-0243

[43] Kassen, M. (2018). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. *Aslib Journal of Information Management*, *70*(5), 518–537. https://doi.org/10.1108/AJIM-11-2017-0250

[44] Kassen, M. (2020). Open data and its peers: Understanding promising harbingers from Nordic Europe. *Aslib Journal of Information Management*, *72*(5), 765–785. https://doi.org/10.1108/AJIM-12-2019-0364

[45] Katsonis, M., & Botros, A. (2015). Digital Government: A Primer and Professional Perspectives: Digital Government. *Australian Journal of Public Administration*, *74*(1), 42–52. https://doi.org/10.1111/1467-8500.12144

[46] Khalifa, M. A., Jennings, M. E., Briscoe, F., Oleszweski, A. M., & Abdi, N. (2014). Racism? Administrative and Community Perspectives in Data-Driven Decision Making: Systemic Perspectives Versus Technical-Rational Perspectives. *Urban Education*, *49*(2), 147–181. https://doi.org/10.1177/0042085913475635

[47] Kim, E. S., Choi, Y., & Byun, J. (2019). Big Data Analytics in Government: Improving Decision Making for R&D Investment in Korean SMEs. *Sustainability*, *12*(1), 202. https://doi.org/10.3390/su12010202

[48] Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews* [Technical Report TR/SE-0401]. https://www.researchgate.net/profile/Barbara-Kitchenham/publication/228756057_Procedures_for_Performing_Systematic_Reviews/links/618cfae961f09877207f8471/Procedures-for-Performing-Systematic-Reviews.pdf

[49] Kitchin, R. (2014a). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

[50] Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

[51] Kitchin, R. (2014c). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

[52] Kitchin, R. (2015). Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society*, *8*(1), 131–136. https://doi.org/10.1093/cjres/rsu027

[53] Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

[54] Kitchin, R., & Dodge, M. (2011). *Code, space: Software and everyday life*. MIT Press.

[55] Kitchin, R., & Lauriault, T. P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. 19.

[56] Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, *2*(1), 6–28. https://doi.org/10.1080/21681376.2014.983149

[57] Klingenberg, C. O., Borges, M. A. V., & Antunes Jr, J. A. V. (2019). Industry 4.0 as a data-driven paradigm: A systematic literature review on technologies. *Journal of Manufacturing Technology Management*, *32*(3), 570–592. https://doi.org/10.1108/JMTM-09-2018-0325

[58] Lacerda, R. T. de O., Ensslin, L., Rolim Ensslin, S., Knoff, L., & Martins Dias Junior, C. (2016). Research Opportunities in Business Process Management and Performance Measurement from a Constructivist View: Research Opportunities in BPM. *Knowledge and Process Management*, *23*(1), 18–30. https://doi.org/10.1002/kpm.1495

[59] Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure: China's Social Credit System as State Surveillance. *Policy & Internet*, *10*(4), 415–453. https://doi.org/10.1002/poi3.183

[60] Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. https://doi.org/10.1016/j.giq.2015.05.006

[61] Lourenço, R. P., Piotrowski, S., & Ingrams, A. (2017). Open data driven public accountability. *Transforming Government: People, Process and Policy*, *11*(1), 42–57. https://doi.org/10.1108/TG-12-2015-0050

[62] Marda, V. (2018). Artificial intelligence policy in India: A framework for engaging the limits of data-driven decision-making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180087. https://doi.org/10.1098/rsta.2018.0087

[63] Matheus, R., & Janssen, M. (2020). A Systematic Literature Study to Unravel Transparency Enabled by Open Government Data: The Window Theory. *Public Performance & Management Review*, *43*(3), 503–534. https://doi.org/10.1080/15309576.2019.1691025

[64] Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, *37*(3), 101284. https://doi.org/10.1016/j.giq.2018.01.006

[65] McBride, K., Aavik, G., Toots, M., Kalvet, T., & Krimmer, R. (2019). How does open government data driven co-creation occur? Six factors and a 'perfect storm'; insights from Chicago's food inspection forecasting model. *Government Information Quarterly*, *36*(1), 88–97. https://doi.org/10.1016/j.giq.2018.11.006

[66] Moro Visconti, R., & Morea, D. (2019). Big Data for the Sustainability of Healthcare Project Financing. *Sustainability*, *11*(13), 3748. https://doi.org/10.3390/su11133748

[67] Parycek, P., Höchtl, J., & Ginner, M. (2014). Open Government Data Implementation Evaluation. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 13–14. https://doi.org/10.4067/S0718-18762014000200007

[68] Peled, A. (2011). When transparency and collaboration collide: The USA Open Data program. *Journal of the American Society for Information Science and Technology*, *62*(11), 2085–2094. https://doi.org/10.1002/asi.21622

[69] Pereira, G. V., Eibl, G., Stylianou, C., Martínez, G., Neophytou, H., & Parycek, P. (2018). The Role of Smart Technologies to Support Citizen Engagement and Decision Making: The SmartGov Case. *International Journal of Electronic Government Research*, *14*(4), 17. https://doi.org/10.4018/IJEGR.2018100101

[70] Phillips-Wren, G., & Hoskisson, A. (2015). An analytical journey towards big data. *Journal of Decision Systems*, *24*(1), 87–102. https://doi.org/10.1080/12460125.2015.994333

[71] Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big Data for Policymaking: Great Expectations, but with Limited Progress?: Big Data for Policymaking. *Policy & Internet*, *10*(3), 347–367. https://doi.org/10.1002/poi3.176

[72] Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

[73] Severo, M., Feredj, A., & Romele, A. (2016). Soft Data and Public Policy: Can Social Media Offer Alternatives to Official Statistics in Urban Policymaking?: Soft Data and Public Policy. *Policy & Internet*, *8*(3), 354–372. https://doi.org/10.1002/poi3.127

[74] Statista. (2022). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025* [Statista]. https://www.statista.com/statistics/871513/worldwide-data-created/

[75] Tenney, M., & Sieber, R. (2016). Data-Driven Participation: Algorithms, Cities, Citizens, and Corporate Control. *Urban Planning*, *1*(2), 101–113. https://doi.org/10.17645/up.v1i2.645

[76] Toufaily, E., Zalan, T., & Dhaou, S. B. (2021). A framework of blockchain technology adoption: An investigation of challenges and expected value. *Information & Management*, *58*(3), 103444. https://doi.org/10.1016/j.im.2021.103444

[77] van Oort, N., Sparing, D., Brands, T., & Goverde, R. M. P. (2015). Data driven improvements in public transport: The Dutch example. *Public Transport*, *7*(3), 369–389. https://doi.org/10.1007/s12469-015-0114-7

[78] Vergara, S. C. (2004). *Projetos e Relatórios de Pesquisa em Administração*. Atlas.

[79] Waheed, H., Hassan, S.-U., Aljohani, N. R., & Wasif, M. (2018). A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology*, *37*(10–11), 941–957. https://doi.org/10.1080/0144929X.2018.1467967

[80] Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: An empirical view from the Citizen's perspective. *Information Systems Frontiers*, *19*(2), 285–300. https://doi.org/10.1007/s10796-016-9679-1

[81] Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, *31*(2), 123–141. https://doi.org/10.1080/02680939.2015.1035758

[82] Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

[83] Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). *Socio-technical Impediments of Open Data*. *10*(2), 17.

[84] Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1,2), 17–33. https://doi.org/10.3233/IP-140329