

Prakoso Bhairawa Putera

Indonesian Institute of Sciences (email: prak001@lipi.go.id)

Rostiena Pasciana

Universitas Garut (email: rostiena_pasciana@fisip.uniga.ac.id)

Submitted: 15 January 2021, Revised: 31 March 2021, Accepted: 8 June 2021

Putera, Prakoso Bhairawa is a Doctoral Student in the Programme of Administrative Sciences (Universitas Padjadjaran); Head of Bureau for Planning and Finance; & Senior Researcher in Policy and Administration in Indonesian Institute of Sciences. He has conducted much research, especially in public policy, STI Policy, Science Policy, Content Analysis, Bibliometrics, Reform & Innovation, and Bibliometrics. In 2014, he followed the NISTEP Fellowship Program: Spring Short Course for STI Policy in Japan, STEPI Fellowship Grant in South Korea, and Young Researchers Leadership Workshop (Hungary, 2019). <http://orcid.org/0000-0002-1474-602>. Email: prak001@lipi.go.id, prakoso19001@mail.unpad.ac.id

Pasciana, Rostiena is a Doctoral Student in the Program of Administrative Sciences (Universitas Padjadjaran); Head of the Quality Assurance System; & Lecturer at the Faculty of Social and Political Science in Garut University. She obtained her Bachelor of Arts from Padjadjaran University (Faculty of Literature; Now Faculty of Arts). Her Master of Science degree is obtained from Garut University, in the Program of Administrative Sciences. She is an LPDP awardee since 2019 until now. She has research and writing interests in Public services, Public Policy, Gender, Children, and Disabilities. <https://orcid.org/0000-0001-7428-8439> Email: rostiena_pasciana@fisip.uniga.ac.id, rostiena19001@mail.unpad.ac.id

Policy & Governance Review
ISSN 2580-4820
Vol. 5, Issue 3, pp. 220-239
DOI: <https://doi.org/10.30589/pgr.v5i3.440>

Big Data for Public Domain:

A bibliometric and visualized study of the scientific discourse during 2000–2020

Abstract

This article aims to investigate the trend of scientific publications under 'big data and policy' research during the last two decades, including the dynamics of the network structure of researchers and the institutions. Bibliometrics is utilized as a tool to reveal the dynamics of scientific discussions that occur through articles, published in international journals indexed/contained in the Scopus database; meanwhile, the analysis visualization is performed by using VOSviewer 1.6.16. The search results indicate that the United States serves as the country of origin for most productive author affiliations in publishing articles, the University of Oxford (United Kingdom) serves as the home institution for most productive author affiliations, and Williamson, B., from the University of Edinburgh (United Kingdom), is considered as the most prolific writer. In addition, the Swiss Sustainability Journal from MDPI is cited as the source for the most widely discussed publication topic in its journals. Further, 'Big Data for Development: A Review of Promises and Challenges' is regarded as the article with the most references. Additionally, the most discussed topics on 'big data and policy' include smart cities, open data, privacy, artificial intelligence, machine learning, and data science.

Keywords:

big data; policy; bibliometrics; VOSviewer; scientific discourse

Introduction

The current public sector has been inseparable from the need for big data (Munne, 2016), and even big data has become a growing consensus among stakeholders, which emphasizes the importance of employing big data in governance worldwide

(Rogge et al., 2017). The government and its tools have massively utilized big data as a form of Smart Government, referring to the effective implementation of ICT in public services. Several examples of the use of big data in the public sector, such as: Open government

and data sharing, Citizen sentiment analysis, Smart city and Internet of things (IoT) applications, Cybersecurity, and several algorithms have embodied tax data and people's online shopping trends, further analyzed and employed into innovations in government services to raise the effectiveness, efficiency, and satisfaction of the community. The big data revolution can present 'a better society' (Azzone, 2018) if big data can improve the quality of decision-making and decision-makers are aware of the potential and opportunities of using big data itself. Public policy today can be understood as a set of actions that affect the solution of a policy problem, namely dissatisfaction with a particular need, request, or opportunity for public intervention (Milano, 2014). The quality can be measured from the capacity to create shared value (Osborne, 2017).

The three dimensions are expected to improve public value (Azzone, 2018), accommodated by the utilization of big data, aiming: 1) to provide a personalized service, 2) to involve the final users in designing and producing a personalized service, and 3) to change the characteristics of the service over time. Further, big data enables decision-making in real time (Höchtel et al., 2016), which has the potential to revise the traditional model in each stage of the policy cycle and to replace it with a continuous evaluation model. Therefore, big data is pivotal, and it must be fulfilled by the government in accommodating each line of governance, which is in accordance with the public domain (Phillips & Higgott, 1999).

In its development, the discussion about the potential, benefits, and numbers of realities from practice in a number of countries provides an immense discourse in the presence of big data, which was previously regarded as a supporter of government administration; however, it gradually plays an important role in prompt decision making (P.B. Putera et al., 2020). The study of big data and the public sector plays a pivotal

role in comprehensively understanding how the scientific discussions are conducted in articles, published in leading international journals (on the Scopus database, the Web of Science (WOS), and a number of other indexers).

However, research that particularly examines trends in big data and policy publications has been inadequate. Recently, research merely reviews the evolution of "data science and big data" research (Raban & Gordon, 2020), Big data in marketing (Amado et al., 2018), interdisciplinary big data research (Hu & Zhang, 2017), and the evolution of big data research (Halevi & Moed, 2012). Therefore, the bibliometric analysis of big data and policies in scientific publications over the last twenty years (2000-2020) is deemed imperative.

The formulations of the problem in this research examine the trend of scientific publications in big data and policy research during the last two decades (2000-2020) and the dynamics of the network structure of researchers, institutions, and research areas in the field of big data and policy.

This study aims to provide quantitative and statistical analysis of publication trends over a wider range of timescales and to explore network structures and trends that have been established in the last 20 years. The utilization of bibliometric analysis is expected to address the gaps in research, involving the main actors such as the authors and institutions, and the countries and areas of research. In addition, this research is expected to provide global information and discourse from scientific publications that have been published in the last twenty years related to big data and policy.

Methods

Data source - This research is regarded as a bibliometric study, engaging the articles related to big data and policy, which were in the Scopus database from 2000 to 2020. The utilization

of Scopus as a data reference is based on the consideration that Scopus indexed publications serve as the standard in scientific publications as widely used in bibliometric analysis (Falagas et al., 2008) offering search facilities on a particular subject and the ability to perform citation analysis. We compared the content coverage and practical utility of PubMed, Scopus, Web of Science, and Google Scholar. The official Web pages of the databases were used to extract information on the range of journals covered, search facilities and restrictions, and update frequency. We used the example of a keyword search to evaluate the usefulness of these databases in biomedical information retrieval and a specific published article to evaluate their utility in performing citation analysis. All databases were practical in use and offered numerous search facilities. PubMed and Google Scholar are accessed for free. The keyword search with PubMed offers optimal update frequency and includes online early articles; other databases can rate articles by number of citations, as an index of importance. For citation analysis, Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy. PubMed remains an optimal tool in biomedical electronic research. Scopus covers a wider journal range, of help both in keyword searching and citation analysis, but it is currently limited to recent articles (published after 1995). The research framework is presented in **Fig. 1**.

Search strategy- The search strategy is utilized to identify publications in the field of public policy by title or abstract. This study specifically utilizes the Scopus database as of December 10, 2020, navigated through: the search options (TITLE-ABS-KEY (“big data”) AND TITLE-ABS-KEY (“policy”) AND PUBYEAR <2021). In addition, to provide a more comprehensive reference, this research limits the subject area into “Social Sciences”, with the source type of: “journal”, “conference proceeding”, “book”, and

“book series”, written only in English (obtaining the 938 articles).

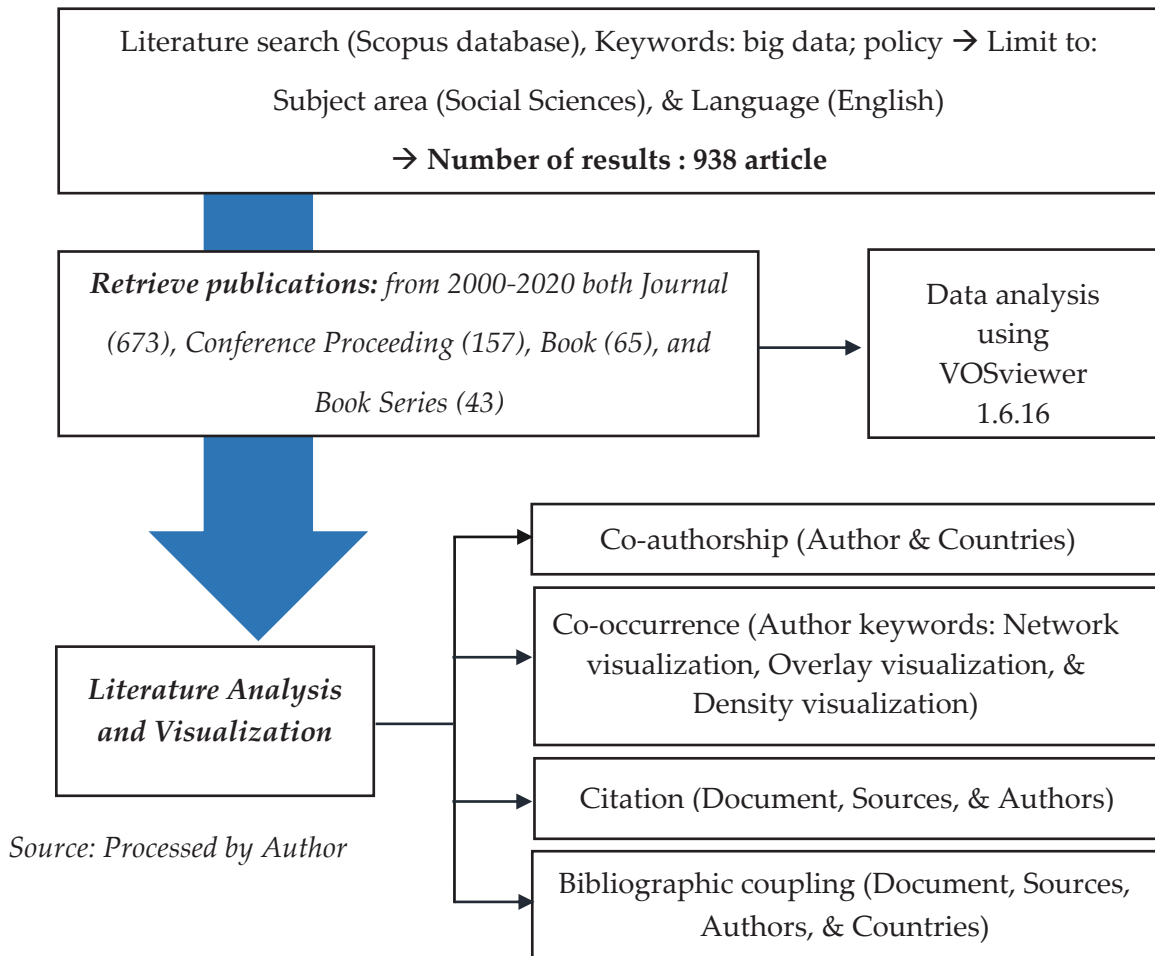
Bibliometric and Visualized analysis

- Bibliometrics is utilized to navigate the development of science and technology through the production of entire scientific literature at a certain level of specialization (Okubo, 1997; Prakoso Bhairawa Putera et al., 2020). Such analysis is valuable to substitute: one country towards another in the global scope; an institution with other institutions within the scope of a country/between countries; and individual researchers/authors in relation to their communities (Prakoso Bhairawa Putera & Rostiena, 2021). Bibliometric data visualization is performed by VOSviewer (analysis software to map analysis based on keywords, authors, countries, and journals) (Eck & Waltman, 2011; Hamidah et al., 2020). Further, bibliometric analysis was generated by employing VOSviewer software version 1.6.16 to visualize, and the map analyzes based on bibliographic data, including the year of publication, overseas affiliates, authors and co-authors, journals, keywords, and citations. This research measures the number of papers written since its inception in 2000 to 2020, by observing the trend of publication on the specific research topic. Additionally, the country and institution of origin of author affiliations was investigated. Hence, this research analyzes the journals that publish big data and policy topics and articles with the highest citation. In the final section, this research presents the results of Cooperation Network Analysis.

Results and Discussion

The development of ‘big data and policy’ publications from 2000-2020 indicates an increase, from 2012 to 2020. The trend analysis (see **Fig. 2**) indicates that during 2001-2011, there were no ‘big data and policy’ publications on the Scopus database.

Figure 1.
Bibliometric Analysis Framework for 'big data and policy'



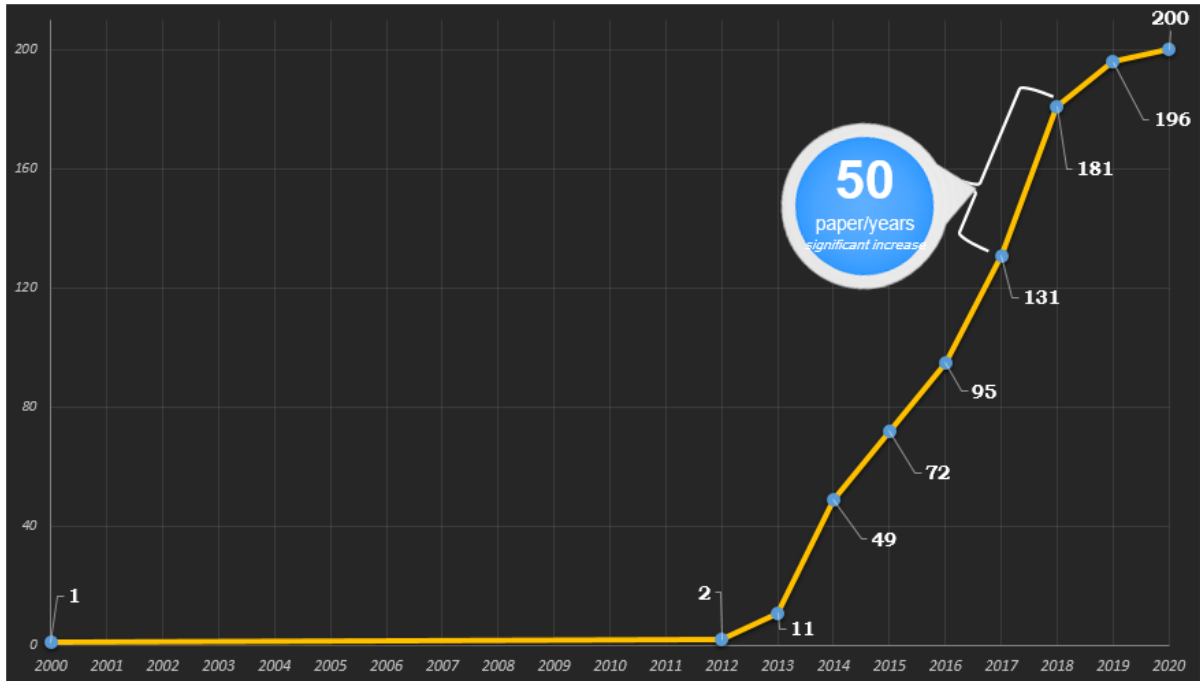
Contribution of Countries/Institutions/ Authors

Based on the bibliometric metadata, it is revealed that there are 78 countries from which the authors' affiliations present the research theme of 'big data and policy', and the Top 10 countries are illustrated in **Fig. 3**. The United States became the most productive country by publishing 277 articles (29.53% of total papers), followed by China with 162 articles (17.27%), and the United Kingdom with 149 articles (15.88%). The results of the bibliometric analysis indicate that there are 9 clusters of author networks based on their affiliated countries. The country cluster is formed by considering the minimum number of documents in each country to 5 articles, in which only 36 out of 83 countries meet the thresholds (see **Fig.4**).

Cluster 1 (in red) indicates a cluster of authors from 8 countries, which include: Belgium, Greece, Netherlands, Poland, Portugal, Saudi Arabia, Spain, and Sweden. In cluster 1, there are 2 countries (the Netherlands and Spain) that publish the most research on the topic of 'big data and policy' (more evidenced by larger nodes than those of other countries). In cluster 2 (in green), there are 6 countries, which include: Austria, Denmark, Japan, Norway, Switzerland, and Taiwan. In this cluster, there are no countries that have a similar publication, indicated by the green nodes which are not larger, except for Hong Kong and Japan with slightly larger nodes.

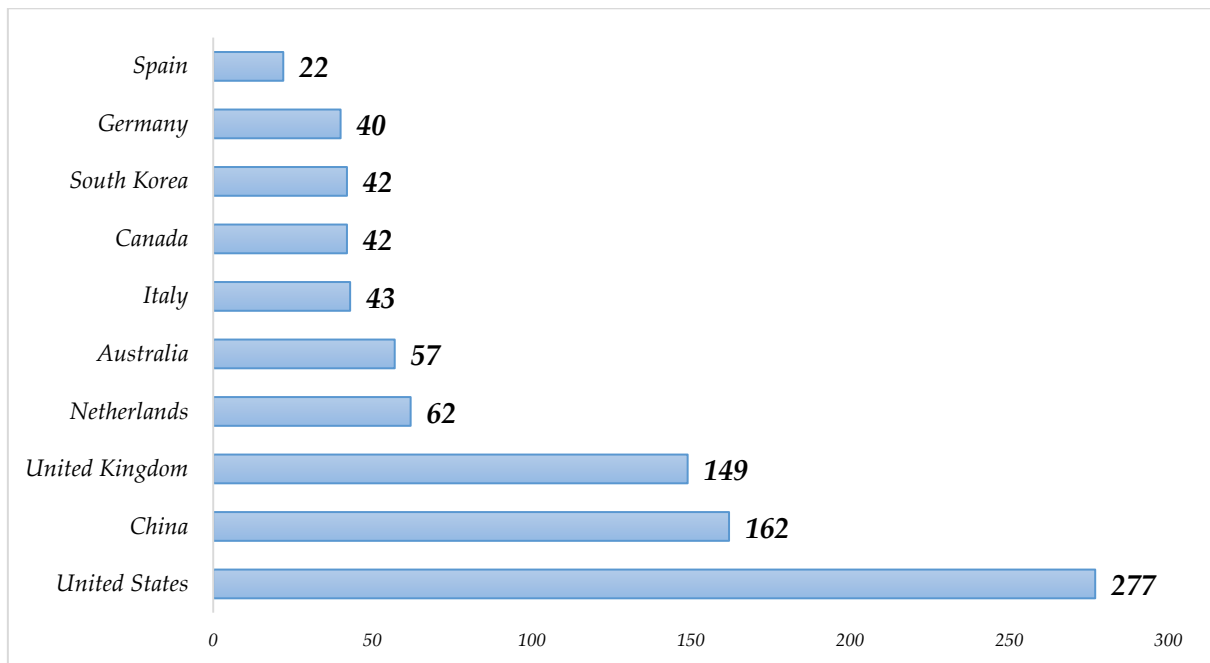
Cluster 3 (in blue) is occupied by the 4 countries, which are: Brazil, France, India, and United Kingdom. This cluster gains more concerns since there is a country (the United Kingdom)

Figure 2.
Publication Trends 'big data and policy'



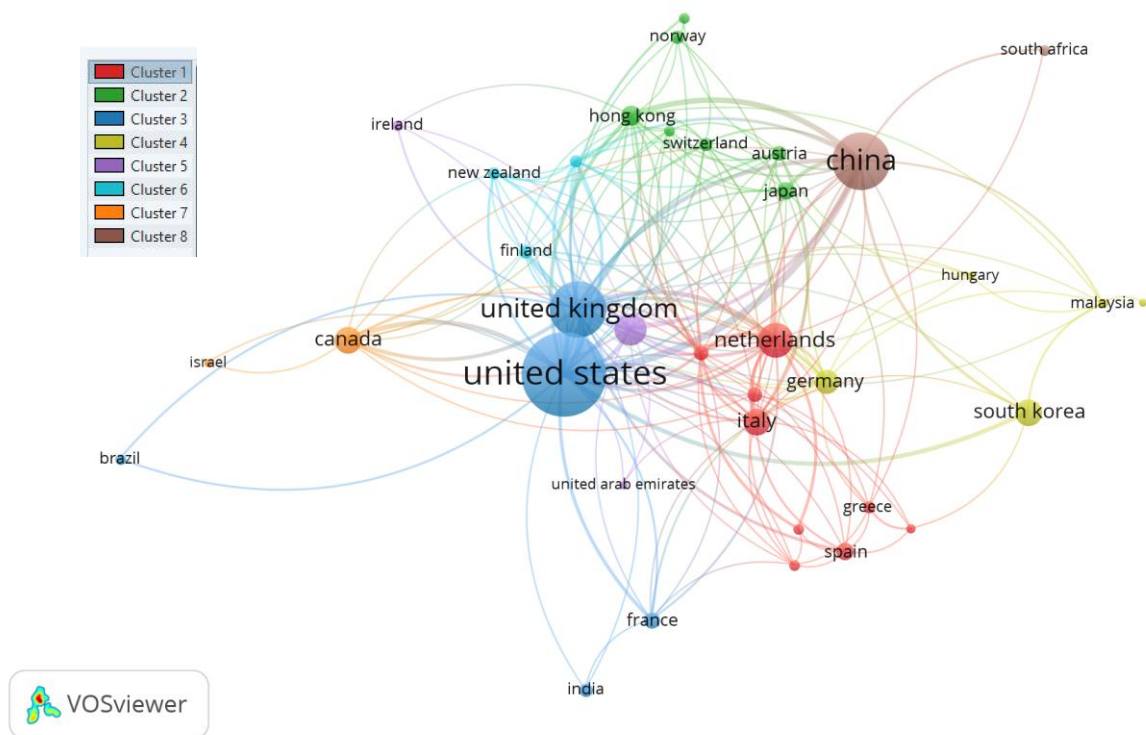
Source: Processed by Author

Figure 3.
Top 10 Most Productive Countries Publishing Research 'big data and policy'



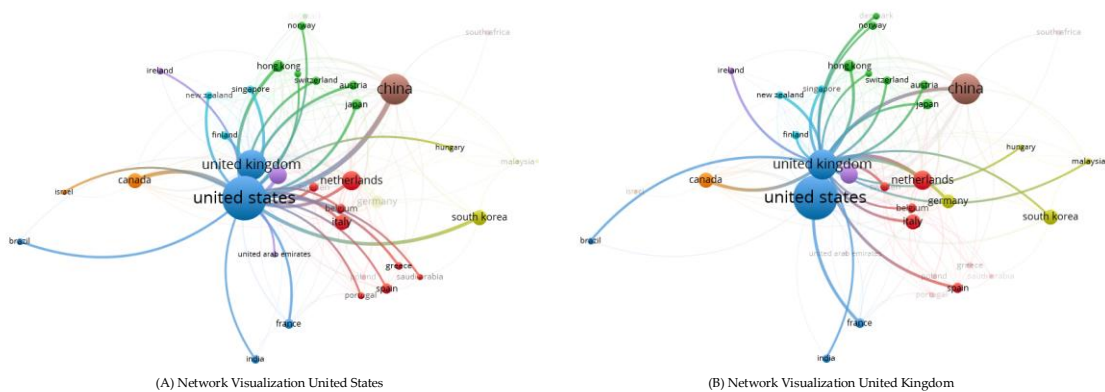
Source: Processed by Author

Figure 4.
Network Visualization of The Most Productive Countries Publishing ‘big data and policy’



Source: Data processed from the VOSviewer 1.6.16 software

Figure 5.
The United States vs United Kingdom Network Visualization Comparison



Source: Data processed from the VOSviewer 1.6.16 software

that produce the most publications in the field of ‘big data and policy’, connected with 27 countries other than the United States (see **Fig. 5**).

Meanwhile, cluster 4 (in yellow) is occupied by the 4 countries, which are: Germany, Hungary, Malaysia, and South Korea. This cluster has 2 larger country nodes for Germany and South Korea. Furthermore, cluster 5 (in purple) is occupied

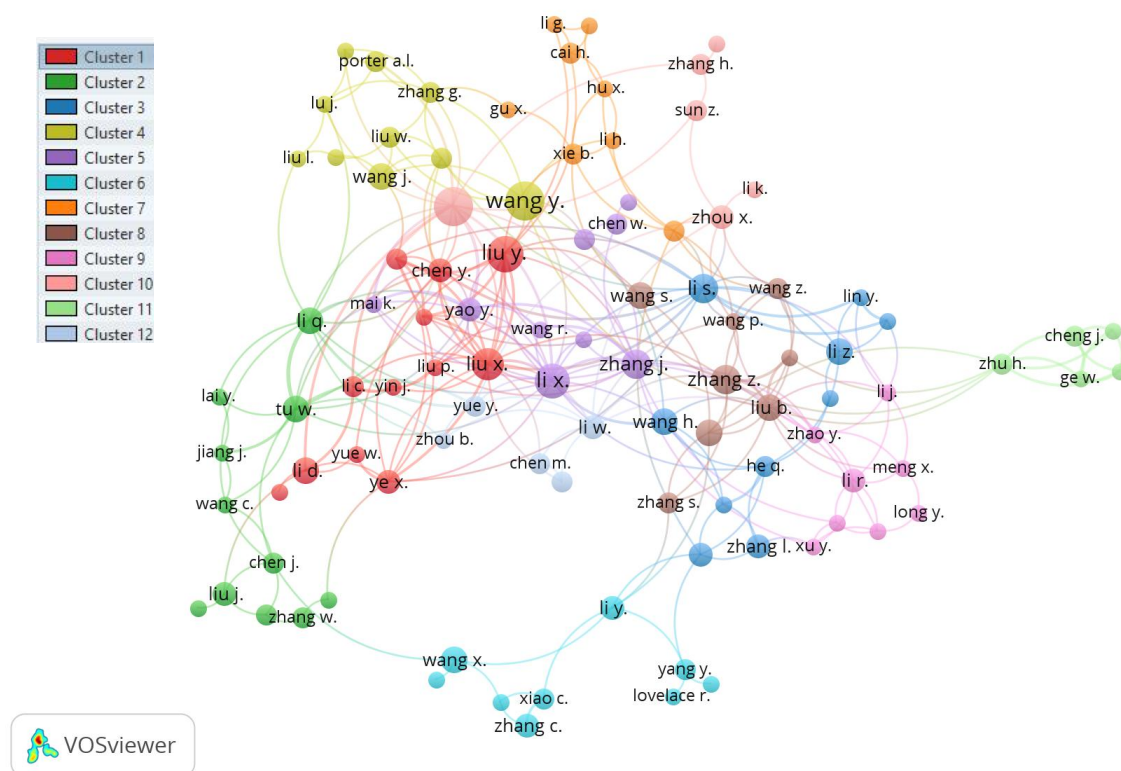
by the 4 countries, which include: Australia, Italy, Russia Federation, and the United Arab Emirates. Cluster 6 (in light blue) is occupied by the 3 countries, which are: Canada, United States, and Israel, depicting the largest nodes in the network (the United States with 29 nodes) connected with other large nodes such as United Kingdom, China, Netherlands, and Australia. Cluster 7 (in orange)

contains Finland, New Zealand, and Singapore. Cluster 8 (in light brown) and Cluster 9 (in pink) each have 2 countries.

In Figure 5A, we provide an example of an explanation that authors with affiliations from the USA have collaborated in writing ‘big data and policy’ articles with authors from affiliations in China, Netherlands, Canada, South Korea, and several other countries. Likewise, writers who come from UK affiliations, shown in Figure 5B, have collaborations in writing articles with authors from affiliations in China, South Korea, Canada, and several other countries. What distinguishes these two examples of collaboration is that the collaboration between the USA and China is more dominant (often co-authoring) compared to the UK and China. This can be seen by the thickness of the relationship line between the USA and China, compared to the USA-China line.

During 2000-2020, 159 institutions became affiliates of authors in the field of ‘big data and policy’ who published scientific papers indexed by Scopus. However, this research presents the 25 most productive institutions crafting the scientific work in the field of ‘big data and policy’ (see **Table 1**). Based on Table 1, it is apparent that the University of Oxford from the United Kingdom serves as the most productive institution with 17 papers (1.81% of the total publications in the field of ‘Big Data and Policy’), followed by 3 institutions, such as Delft University of Technology from the Netherlands, Chinese Academy of Sciences from China, and Leiden University from the Netherlands, which have 13 articles (1.39%). In addition, the fifteen institutions were contributed from five countries, which are: 1) China (6 institutions), 2) United States (6 institutions), 3) United Kingdom (4 institutions), 4) Netherlands (4 institutions), and 5) Australia (2 institutions).

Figure 6.
Network Visualization of Authors Collaborating to Produce Articles Entitled ‘big data and policy’



Source: Data processed from the VOSviewer 1.6.16 software

This finding is in line with previous findings that these five countries hence serve as countries with the most publications in the field of ‘Big Data and Policy’.

Based on the data of the most productive authors producing ‘big data and policy’ articles, we can see that over the past 20 years, there have been more than 2027 authors, either individually or collaborating with fellow authors, that published Scopus indexed articles under the topic of ‘Big Data and Policy,’ and the 25 authors with the most publications as depicted in Table 2.

Table 2 shows that the authors who produced the most ‘big data and policy’ articles came from Netherlands affiliations (6 authors), followed by authors from UK affiliations (5 authors), followed by

authors with affiliations from China USA respectively. Each country has 4 authors, two authors each affiliated with Australia and Norway, and from country affiliations, Denmark and Belgium, one author each. Let’s look at the distribution of continents. Writers with affiliations in European countries are the most productive in producing articles, followed by writers with affiliations from Asian, American, and Australian countries. Let’s look at the distribution of continents. Writers with associations in European countries are the most productive in producing articles, followed by writers with affiliations from Asian, American, and Australian countries.

Williamson, B. from the University of Edinburgh (United Kingdom) is recorded as

Table 1.
Top 25 Author Affiliate Institutions Researching “Big Data and Policy”

Rank	Institutions	Country	Number of Paper	Citations	Quality of Paper	% of Total Paper
1 st	University of Oxford	United Kingdom	17	307	18.06	1.81
2 nd	Delft University of Technology	Netherlands	13	150	11.54	1.39
3 rd	Chinese Academy of Sciences	China	13	56	4.31	1.39
4 th	Leiden University	Netherlands	13	104	8.00	1.39
5 th	Sun Yat-Sen University	China	12	222	18.50	1.28
6 th	New York University	United States	11	259	23.55	1.17
7 th	Wuhan University	China	11	113	10.27	1.17
8 th	Ministry of Education China	China	10	51	5.10	1.07
9 th	Harvard University	United States	10	54	5.40	1.07
10 th	University of Leeds	United Kingdom	10	47	4.70	1.07
11 th	The University of Edinburgh	United Kingdom	10	134	13.40	1.07
12 th	The University of Queensland	Australia	10	209	20.90	1.07
13 th	University of Cambridge	United States	10	69	6.90	1.07
14 th	Oxford Social Sciences Division	United Kingdom	10	233	23.30	1.07
15 th	Shenzhen University	China	9	91	10.11	0.96
16 th	Universiteit van Amsterdam	Netherlands	9	231	25.67	0.96
17 th	The University of Hong Kong	Hong Kong	9	84	9.33	0.96
18 th	Peking University	China	9	108	12.00	0.96
19 th	University of Toronto	Canada	9	67	7.44	0.96
20 th	Tilburg University	Netherlands	9	51	5.67	0.96
21 th	National University of Singapore	Singapore	9	95	10.44	0.96
22 th	UNSW Sydney	Australia	9	100	11.11	0.96
23 th	Pennsylvania State University	United States	8	94	11.75	0.85
24 th	Arizona State University	United States	8	185	23.13	0.85
25 th	The University of Texas at Austin	United States	8	208	26.00	0.85

Source: Processed by Author from Scopus database

Table 2.
Top 25 Authors of “Big Data and Policy” Based on the number of Published Papers

Rank	Name	Institutions	Country	Number of Paper	Citations	Quality of Paper	% of Total Paper
1 st	Williamson, B.	University of Edinburgh	United Kingdom	7	251	35.86	0.75
2 nd	Bertino, E.	Purdue University, West Lafayette	United States	5	21	4.20	0.53
3 rd	Pettit, C.	University of New South Wales	Australia	5	22	4.40	0.53
4 th	Taylor, L.	Tilburg Institute for Law, Technology and Society (TILT), Tilburg University,	Netherlands	5	148	29.60	0.53
5 th	Tu, W.	Shenzhen University	China	5	66	13.20	0.53
6 th	Bibri, S.E.	The Norwegian University of Science and Technology	Norway	4	9	2.25	0.43
7 th	Giest, S.	Institute of Public Administration, Leiden University, Den Haag	Netherlands	4	53	13.25	0.43
8 th	Janssen, M.	Delft University of Technology, Faculty of Technology, Policy and Management	Netherlands	4	81	20.25	0.43
9 th	Karafili, E.	Imperial College London	United Kingdom	4	4	1.00	0.43
10 th	Kontokosta, C.E.	Marron Institute of Urban Management, New York University	United States	4	49	12.25	0.43
11 th	Kourtit, K.	JADS (Jheronimus Academy of Data Science), Hertogenbosch	Netherlands	4	49	12.25	0.43
12 th	Li, Q.	Shenzhen University	China	4	66	16.50	0.43
13 th	Liu, X.	Sun Yat-sen University, Guangzhou	China	4	124	31.00	0.43
14 th	Schroeder, R.	University of Oxford	United Kingdom	4	83	20.75	0.43
15 th	Yao, Y.	China University of Geosciences, Wuhan	China	4	102	25.50	0.43
16 th	Bibri, S.E.	Norges teknisk-naturvitenskapelige universitet, Trondheim	Norway	3	8	2.67	0.32
17 th	Custers, B.	Leiden University, Center for Law and Digital Technologies, Leiden	Netherlands	3	44	14.67	0.32
18 th	Desouza, K.C.	QUT Business School, Brisbane	Australia	3	65	21.67	0.32
19 th	Hoeyer, K.	Københavns Universitet, Department of Public Health, Copenhagen	Denmark	3	20	6.67	0.32
20 th	Margetts, H.	Oxford Social Sciences Division, Oxford	United Kingdom	3	94	31.33	0.32
21 th	Mureddu, F.	The Lisbon Council for Economic Competitiveness and Social Renewal asbl, Brussels	Belgium	3	18	6.00	0.32
22 th	Nijkamp, P.	Vrije Universiteit Amsterdam, Amsterdam	Netherlands	3	56	18.67	0.32
23 th	Porter, A.L.	Georgia Institute of Technology, Atlanta	United States	3	5	1.67	0.32
24 th	Williams, C.	Defence Science and Technology Laboratory, Salisbury	United Kingdom	3	20	6.67	0.32
25 th	Ye, X.	Texas A&M University, Department of Landscape Architecture and Urban Planning, College Station	United States	3	52	17.33	0.32

Source: Processed by Author from Scopus database

Table 3.
Top 25 Publication Sources of 'big data and policy' based on number of publications

Rank	Source Title/subject area of the journal	Publishers	Type of Source	Scopus quartiles	Number of Paper	Citations	Quality of Paper	% of Total Paper
1 st	Sustainability Switzerland / <i>Social Sciences: Geography, Planning, and Development</i>	Multidisciplinary Digital Publishing Institute (MDPI)	Journal	Q1	38	385	10,13	4,16%
2 nd	Policy and Internet / <i>Social Sciences: Public Administration</i>	Wiley-Blackwell	Journal	Q1	22	647	29,41	2,35%
3 rd	2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/ SCALCOM/ UIC/ATC/ CBDCOM/IOP/SCI 2017 - Conference Proceedings / <i>Computer Science: Computer Networks and Communications</i>	IEEE	Proceedings	-	21	36	1,86	2,24%
4 th	Big Data and Society / <i>Social Sciences: Communication</i>	SAGE	Journal	Q1	21	662	31,52	2,24%
5 th	Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/ BDCloud/ SustainCom/ SocialCom 2019 / <i>Computer Science: Computer Networks and Communications</i>	IEEE	Proceedings	-	17	1	0,06	1,82%
6 th	Cities / <i>Social Sciences: Development</i>	Elsevier Ltd	Journal	Q1	13	234	18,00	1,39%
7 th	Government Information Quarterly / <i>Social Sciences: Law</i>	Elsevier Ltd	Journal	Q1	12	214	17,83	1,28%
8 th	Computer Law and Security Review / <i>Social Sciences: Law</i>	Elsevier Ltd	Journal	Q1	11	93	8,45	1,17%
9 th	Computers Environment and Urban Systems / <i>Social Sciences: Geography, Planning, and Development</i>	Elsevier Ltd	Journal	Q1	10	112	11,20	1,07%

continues to page 230

continued from page 229

10 th	Proceedings - 2016 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2016 / Computer Science: Computer Networks and Communications	IEEE	Proceedings	-	10	6	0,60	1,07%
11 th	Review of Policy Research / <i>Social Sciences: Public Administration</i>	Blackwell Publishing	Journal	Q1	10	234	23,40	1,07%
12 th	International Archives of The Photogrammetry Remote Sensing and Spatial Information Sciences ISPRS Archives / Computer Science: Computer Networks and Communications	International Society for Photogrammetry and Remote Sensing	Proceedings	-	9	9	1,00	0,96%
13 th	Land Use Policy / <i>Social Sciences: Geography, Planning and Development</i>	Elsevier Ltd	Journal	Q1	9	44	4,89	0,96%
14 th	Proceedings 2020 International Conference on Big Data and Informatization Education / <i>Computer Science: Computer Networks and Communications</i>	IEEE	Proceedings	-	9	1	0,11	0,96%
15 th	Public Administration and Information Technology / <i>Social Sciences: Public Administration</i>	Springer	Journal	Q2	9	40	4,44	0,96%
16 th	Proceedings 2015 IEEE International Conference On Smart City Smartcity / <i>Computer Science: Computer Networks and Communications</i>	IEEE	Proceedings	-	8	34	4,25	0,85%
17 th	Telecommunications Policy / <i>Social Sciences: Human Factors and Ergonomics</i>	Elsevier	Journal	Q1	8	141	17,63	0,85%
18 th	International Perspectives on Education and Society / <i>Social Sciences: Sociology and Political Science</i>	Emerald	Journal	Q3	7	1	0,14	0,75%
19 th	Journal of Policy Analysis and Management / <i>Social Sciences: Public Administration</i>	Wiley-Blackwell	Journal	Q1	7	71	10,14	0,75%
20 th	Journal of Transport Geography / <i>Social Sciences: Geography, Planning and Development</i>	Elsevier	Journal	Q1	7	97	13,86	0,75%
21 th	ISPRS International Journal of Geo Information / <i>Social Sciences: Geography, Planning and Development</i>	Multidisciplinary Digital Publishing Institute (MDPI)	Journal	Q1	6	45	7,50	0,64%

continues to page 231

continued from page 230

22 th	Springer Geography / <i>Social Sciences: Geography, Planning and Development</i>	Springer Nature	Journal	Q2	6	37	6,17	0,64%
23 th	Transforming Government People Process and Policy <i>/ Social Sciences: Public Administration</i>	Emerald	Journal	Q1	6	10	1,67	0,64%
24 th	Transportation Research Part C Emerging Technologies / <i>Social Sciences: Transportation</i>	Elsevier	Journal	Q1	6	338	56,33	0,64%
25 th	Academic Medicine / <i>Social Sciences: Education</i>	Wolters Kluwer Health	Journal	Q1	5	67	13,40	0,53%

Source: Processed by Author from Scopus database

having the most publications as well as the most citations (251 citations), followed by four other authors (Bertino, E from Purdue University, West Lafayette, United States; Pettit, C from University of New South Wales, Australia; Taylor, L from Tilburg Institute for Law, Technology and Society (TILT), Tilburg University, Netherlands; and Tu, W from Shenzhen University, China) who have five articles. When viewed from the Quality of paper, measured from the total citation / total paper produced, there are differences in the arrangement, where Liu occupies the second position. X from Sun Yat-sen University, Guangzhou, China, with four publications having 124 citations.

Distribution of Journals and Highly Cited Articles

During 2000-2020, there were 939 publications published on the topic 'big data and policy' based on the Scopus database. Most of them were published in the Journal (71.78%), Conference Proceedings (16.40%), Book (6.92%), and Book Series (4.58%). The visualization of the publication sources 'Big Data and Policy' forms six clusters (see Figure 7). These results are in accordance with Table 3. Based on **Table 3**, it is apparent that the 25 sources of 'big data and policy' publications were based on the number of publications published. The journal entitled 'Sustainability Switzerland' from MDPI

publisher was recorded as the publication source that published the most articles (38 articles or 4.16% of the total papers published on this topic, with a total of 385 citations, with a quality of paper/citations of 10.13) under the 'big data and policy' topic during 2000-2020. This finding means that every published paper is referred to 10 times. The second position is occupied by the journal of 'Policy and Internet' from the Wiley-Blackwell publisher, comprising 22 articles with 647 citations. Lastly, the third place is occupied by the proceedings of the '2017 IEEE Smartworld Ubiquitous Intelligence and Computing Conference Proceedings' from the IEEE publisher, comprising 21 papers on the topic of 'big data and policy'.

Based on the mapping results and shown in Table 3, of the 25 publication sources that mainly publish 'big data and policy' articles, most of the articles are published in journals (76%), and 24% come from the Proceedings. Journals with Scopus quartiles Q1 publish the most 'big data and policy' articles (as many as 84.21% or 16 journals), Q2 (as many as 10.53% or only two journals), and Q3 (as many as 5.36% or one journal only). This indicates that the article 'big data and policy' is a topic of interest published in high-reputation journals (Q1).

In Table 3, the subject areas of the journal are in the fields of Social Sciences: Geography,

Planning and Development, and Computer Science: Computer Networks and Communications (24% each), followed by Social Sciences: Public Administration (20%), Social Sciences: Law (8%), and six other fields, namely Social Sciences: Communication, Social Sciences: Development, Social Sciences: Human Factors and Ergonomics, Social Sciences: Sociology and Political Science, Social Sciences: Transportation, Social Sciences: Education (1% each). This indicates that 'big data and policy' is discussed in the field of Public Administration and has become a concern of other areas in the social sciences and computer science families.

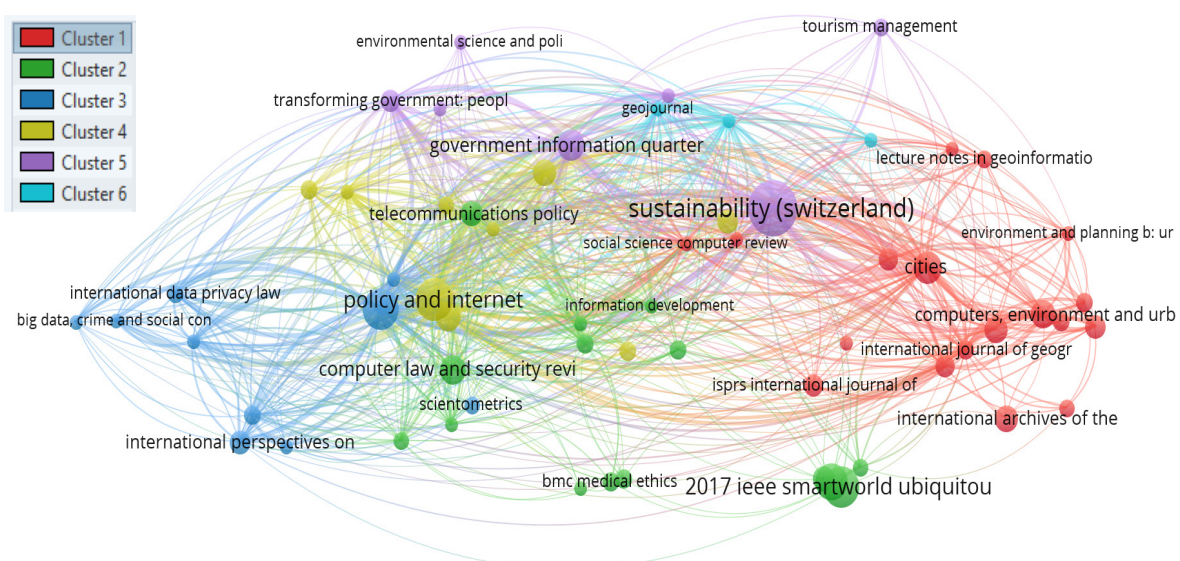
Meanwhile, of the 939 published 'big data and policy' articles, it had 9,160 citations. **Table 4** indicates the 25 most cited publication titles. The article entitled 'Big Data for Development: A Review of Promises and Challenges' from 'Development Policy Review, 34 (1), pp. 135-174' (Hilbert, 2016) is recorded as the article with the most citations, containing 209 citations with an average annual citation (C/Y) of 52.25. The second place is occupied by the article entitled 'The path most traveled:

Travel demand estimation using big data resources' from 'Transportation Research Part C: Emerging Technologies, 58, pp. 162-177' (Toole et al., 2015), which has 201 citations with an average annual citation of 40.20. The third position is occupied by the article entitled 'Flood prediction using machine learning models: Literature review' from the 'Water (Switzerland), 10(11), 1536' (Mosavi et al., 2018), which has been cited as many as 195 citations with an average annual citation of 97.50. In addition, **Table 4** presents that the top 25 articles with the most citations are from the Wiley-Blackwell and MDPI publishers. **Fig. 8** indicates that the network visualization of the article with a minimum of 5 per article citation has formed the 11 network clusters.

Keyword analysis

This analysis is regarded as co-occurrence, which aims to present a visualization of the network between keywords. Based on **Fig. 9**, it is apparent that the analysis using the 'author keywords' of 939 'big data and policy' articles contains 2,879 keywords, in which this research records a minimum number of occurrences from

Figure 7.
Publication Source Network of 'Big Data and Policy'



Source: Data processed from the VOSviewer 1.6.16 software

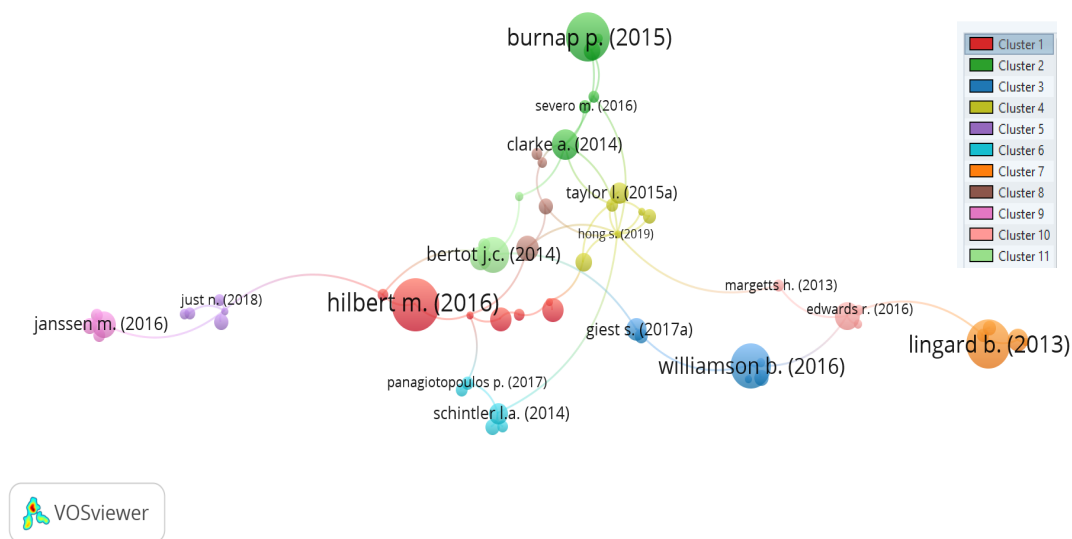
Table 4.
Top 15 Articles on 'big data and policy' with the Most Citations

Rank	Title	Year	Source Title	Publishers	Citations	C/Y
1 st	Big Data for Development: A Review of Promises and Challenges (Hilbert, 2016)	2016	Development Policy Review	Blackwell Publishing Ltd	209	52.25
2 nd	The path most traveled: Travel demand estimation using big data resources (Toole et al., 2015)	2015	Transportation Research Part C: Emerging Technologies	Elsevier Ltd	201	40.20
3 rd	Flood prediction using machine learning models: Literature review (Mosavi et al., 2018)	2018	Water (Switzerland)	MDPI AG	195	97.50
4 th	Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making (Burnap & Williams, 2015)	2015	Policy and Internet	Wiley-Blackwell Publishing Ltd	185	37.00
5 th	Testing regimes, accountabilities, and education policy: Commensurate global and national developments (Lingard et al., 2013)	2013	Journal of Education Policy		185	26.43
6 th	Accountable algorithms (Kroll et al., 2017)	2017	University of Pennsylvania Law Review	University of Pennsylvania Law School	184	61.33
7 th	Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments (Williamson, 2016)	2016	Journal of Education Policy	Routledge	151	37.75
8 th	Big Data Surveillance: The Case of Policing (Brayne, 2017)	2017	American Sociological Review	American Sociological Association	136	45.33
9 th	Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet (Cai et al., 2014)	2014	Transportation Research Part D: Transport and Environment	Elsevier Ltd	135	22.50
10 th	Big data, bigger dilemmas: A critical review (Ekbja et al., 2015)	2015	Journal of the Association for Information Science and Technology	John Wiley and Sons Inc.	131	26.20
11 th	Industry 4.0 and sustainability implications: A scenario-based analysis of the impacts and challenges (Bonilla et al., 2018)	2018	Sustainability (Switzerland)	MDPI AG	113	56.50
12 th	The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making (Zarsky, 2016)	2016	Science Technology and Human Values	SAGE Publications Inc.	108	27.00
13 th	Smart cities with big data: Reference models, challenges, and considerations (Lim et al., 2018)	2018	Cities	Elsevier Ltd	101	50.50
14 th	Big data, open government, and e-government: Issues, policies, and recommendations (Bertot et al., 2014)	2014	Information Polity	IOS Press	100	16.67
15 th	Grand challenges in altmetrics: heterogeneity, data quality, and dependencies (Haustein, 2016)	2016	Scientometrics	Springer Netherlands	99	24.75
16 th	Civic Hackathons: Innovation, procurement, or civic engagement? (Johnson & Robinson, 2014)	2014	Review of Policy Research	Blackwell Publishing Inc.	99	16.50
17 th	On big data, artificial intelligence, and smart cities (Allam & Dhunny, 2019)	2019	Cities	Elsevier Ltd	98	98.00

18 th	The digital skin of cities: Urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data (Rabari & Storper, 2015)	2015	Cambridge Journal of Regions, Economy, and Society	Oxford University Press	95	19.00
19 th	The datafication of health (Ruckenstein & Schüll, 2017)	2017	Annual Review of Anthropology	Annual Reviews Inc.	92	30.67
20 th	Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization (Fu et al., 2018)	2018	Sustainability (Switzerland)	MDPI AG	90	45.00
21 th	Developing the role of big data and analytics in health professional education (Ellaway et al., 2014)	2014	Medical Teacher	Taylor & Francis	81	13.50
22 th	Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data (Yao et al., 2017)	2017	International Journal of Geographical Information Science	Taylor and Francis Ltd.	76	25.33
23 th	Communication dynamics in Twitter during political campaigns: The case of the 2011 Spanish national election (Aragón et al., 2013)	2013	Policy and Internet	Wiley-Blackwell	73	10.43
24 th	An empirical study on travel patterns of internet-based ride-sharing (Dong et al., 2018)	2018	Transportation Research Part C: Emerging Technologies	Elsevier Ltd	70	35.00
25 th	Governments and citizens getting to know each other. Open, closed, and big data in public management reform (Clarke & Margetts, 2014)	2014	Policy and Internet	Wiley-Blackwell Publishing Ltd	70	11.67

Source: Processed by Author from Scopus database

Figure 8.
Network Visualization from the Article Entitled 'Big Data and Policy'



Source: Data processed from the VOSviewer 1.6.16 software

a keyword with 5 repetition keywords. The result obtains 87 keywords that meet the limitation and form 9 clusters (see **Fig.9**).

Based on **Fig. 9**, it is apparent that the smart city research topic has the largest nodes in cluster 1 (in red), indicating that the topic is predominantly discussed around 2019 (see **Fig.10**). In addition,

- on *Management and Business Economics*, 24(1), 1–7. <https://doi.org/10.1016/j.iedeen.2017.06.002>
- Aragón, P., Kappler, K. E., Kaltenbrunner, A., Laniado, D., & Volkovich, Y. (2013). Communication dynamics in twitter during political campaigns: The case of the 2011 Spanish national election. *Policy and Internet*, 5(2), 183–206. <https://doi.org/10.1002/1944-2866.POI327>
- Azzone, G. (2018). Big data and public policies: Opportunities and challenges. *Statistics and Probability Letters*, 136, 116–120. <https://doi.org/10.1016/j.spl.2018.02.022>
- Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19(1–2), 5–16. <https://doi.org/10.3233/IP-140328>
- Bonilla, S. H., Silva, H. R. O., da Silva, M. T., Gonçalves, R. F., & Sacomano, J. B. (2018). Industry 4.0 and sustainability implications: A scenario-based analysis of the impacts and challenges. *Sustainability (Switzerland)*, 10(10). <https://doi.org/10.3390/su10103740>
- Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review*, 82(5), 977–1008. <https://doi.org/10.1177/0003122417725865>
- Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy and Internet*, 7(2), 223–242. <https://doi.org/10.1002/poi3.85>
- Cai, H., Jia, X., Chiu, A. S. F., Hu, X., & Xu, M. (2014). Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D: Transport and Environment*, 33, 39–46. <https://doi.org/10.1016/j.trd.2014.09.003>
- Clarke, A., & Margetts, H. (2014). Governments and citizens getting to know each other? open, closed, and big data in public management reform. *Policy and Internet*, 6(4), 393–417. <https://doi.org/10.1002/1944-2866.POI377>
- Dong, Y., Wang, S., Li, L., & Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing. *Transportation Research Part C: Emerging Technologies*, 86, 1–22. <https://doi.org/10.1016/j.trc.2017.10.022>
- Eck, N. J. Van, & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter*, 7(3), 50–54.
- Ekbria, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(8), 1523–1545. <https://doi.org/10.1002/asi.23294>
- Ellaway, R. H., Pusic, M. V, Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, 36(3), 216–222. <https://doi.org/10.3109/0142159X.2014.874553>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492lsf>
- Fu, H., Li, Z., Liu, Z., & Wang, Z. (2018). Research on big data digging of hot topics about recycled water use on micro-blog based on particle swarm optimization. *Sustainability (Switzerland)*, 10(7). <https://doi.org/10.3390/su10072488>
- Halevi, G., & Moed, H. F. (2012). The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. *Bibliometrics*, 30. <https://www.researchtrends.com/>

issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/

- Hamidah, I., Sriyono, & Hudha, M. N. (2020). A Bibliometric Analysis of Covid-19 Research using VOSviewer. *Indonesia Journal Of Science and Technology*, 5(2), 209–216.
- Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413–423. <https://doi.org/10.1007/s11192-016-1910-9>
- Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, 34(1), 135–174. <https://doi.org/10.1111/dpr.12142>
- Höchtel, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, 26(1–2), 147–169. <https://doi.org/10.1080/10919392.2015.1125187>
- Hu, J., & Zhang, Y. (2017). Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. *Scientometrics*, 112(1), 91–109. <https://doi.org/10.1007/s11192-017-2383-1>
- Johnson, P., & Robinson, P. (2014). Civic Hackathons: Innovation, procurement, or civic engagement? *Review of Policy Research*, 31(4), 349–357. <https://doi.org/10.1111/ropr.12074>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85012962597&partnerID=40&md5=648224f63df96b351aba9b8513682b88>
- Lim, C., Kim, K.-J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, 82, 86–99. <https://doi.org/10.1016/j.cities.2018.04.011>
- Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Education Policy*, 28(5), 539–556. <https://doi.org/10.1080/02680939.2013.820042>
- Milano, P. (2014). Understanding policy decisions. In *Springer Briefs in Applied Sciences and Technology* (Vol. 14, Issue January 2014). https://doi.org/10.1007/978-3-319-02520-9_1
- Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water (Switzerland)*, 10(11). <https://doi.org/10.3390/w10111536>
- Munne, R. (2016). Big Data in the Public Sector. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 195–208). Springer Nature. <https://doi.org/10.1007/978-3-319-21569-3>
- Okubo, Y. (1997). *Bibliometric Indicators and Analysis of Research Systems: Methods and Examples* (OECD Science, Technology and Industry Working Papers 1997/01). <https://doi.org/https://dx.doi.org/10.1787/208277770603>
- Phillips, N., & Higgott, R. (1999). *Global Governance and the Public Domain: Collective Goods in a “Post-Washington Consensus” Era* (No. 47; 99, Issue 47).
- Putera, P.B., Manik, L. P., Rianto, Y., Sari, A. A., & Sadikin, R. (2020). How indonesia uses big data “indonesian one data” for the future of policy making. *International Journal of Advanced Science and Technology*, 29(5).
- Putera, Prakoso Bhairawa, & Rostiena, P. (2021). Identity, Intellectual and Paradigm

- Dialogue: India and Indonesia Public Administration Research in Global Order. *Library Philosophy and Practice*, 4885, 1–16. <https://digitalcommons.unl.edu/libphilprac/4885>
- Putera, Prakoso Bhairawa, Suryanto, S., Ningrum, S., & Widianingsih, I. (2020). A bibliometric analysis of articles on innovation systems in Scopus journals written by authors from Indonesia, Singapore, and Malaysia. *Science Editing*, 7(2), 177–183. <https://doi.org/10.6087/KCSE.214>
- Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, 122(3), 1563–1581. <https://doi.org/10.1007/s11192-020-03371-2>
- Rabari, C., & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, 8(1), 27–42. <https://doi.org/10.1093/cjres/rsu021>
- Rogge, N., Agasisti, T., & De Witte, K. (2017). Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. *Public Policy and Administration*, 32(4), 263–281. <https://doi.org/10.1177/0952076716687355>
- Ruckenstein, M., & Schüll, N. D. (2017). The datafication of health. In *Annual Review of Anthropology* (Vol. 46, pp. 261–278). Annual Reviews Inc. <https://doi.org/10.1146/annurev-anthro-102116-041244>
- Secchi, P. (2018). On the role of statistics in the era of big data: A call for a debate. *Statistics and Probability Letters*, 136, 10–14. <https://doi.org/10.1016/j.spl.2018.02.041>
- Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58, 162–177. <https://doi.org/10.1016/j.trc.2015.04.022>
- 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>
- Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., & Zhang, Y. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science*, 31(6), 1220–1244. <https://doi.org/10.1080/13658816.2017.1290252>
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science Technology and Human Values*, 41(1), 118–132. <https://doi.org/10.1177/0162243915605575>