Mapping the Research on Open Government Data: a bibliometric analysis

Dr. Sudip Banerjee

Librarian, City College of Commerce and Business Administration, Kolkata

Susmita Sarkar Basu

Librarian, Bankim Sardar College, Tangrakhali, South 24 Parganas

Abstract

The purpose of this study is to analyse the publication outputs of open government data research using bibliometric indicators and statistical tools. For this study, data spanning 2011 to 2022 were gathered from the Scopus database, examining aspects like year-wise distribution, authorship patterns, collaborations, citation impact, document types, prolific authors, core journals, and leading countries. The research output on open government data exhibited consistent growth, totaling 1129 publications from 2011 to 2022. Collaboration was prominent, with the USA leading in research contributions with 132 publications. Despite a majority receiving citations, 27.19% had none. The dataset did not strictly conform to Bradford's law of scattering. The study suggests a promising future for open government data research, offering insights into future trends and identifying key themes in the field.

Keywords: Bibliometric, Data, Government data, Open data, Open government data, Research mapping

1. Introduction

In recent years, several open data movements started around the world with two major aims, one is transparency and another is data reuse (Attard, Orlandi, Scerri, & Auer, 2015).The Rio Declaration on Environment and Development (1992) of the United Nations, emphasised the importance of data availability to the public. The Organisation for Economic Cooperation and Development defines Open Government Data (OGD) as "a philosophy- and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all" (Open Government Data - OECD, n.d.). Open government is almost synonymous with e-government, but they are different in nature. While e-government focuses on delivering services, open government goes beyond, providing extensive data access and transparent accounts to society (Junior, De Almeida, Segundo, Da Costa, & El-Aouar, 2020). National Data Sharing and Accessibility Policy(2012) identified benefits of OGD like maximum use, maximum integration, avoiding duplication, and better decision-making etc. Governments worldwide leverage portals like Data.gov, data.gov.uk, and others to foster transparency ("Data.gov Home," n.d.; "Find Open Data - data.gov.uk," n.d.; & "Home," n.d.).

Research on open government data is crucial for enhancing data quality and standards, ensuring accuracy, reliability, and universal accessibility. Furthermore, such research guides the utilisation of open data for positive social, economic, and political impacts. Measuring these research outcomes is hence important. This study aims to evaluate the global publication landscape on OGD through bibliometric analysis, examining researchers' contributions and progress as highlighted by Broadus (2017) by expanding the definition scope of bibliometric.

2. Literature review

The landscape of open government data (OGD) has witnessed significant growth and evolution over the past decade. This literature review aims to provide a thematic analysis of key contributions in OGD research, offering insights into the diverse facets that have shaped this field.

The foundation of OGD research is rooted in the eight principles outlined by the Open Government Working Group in 2007, emphasising data openness, completeness, timeliness, accessibility, non-discrimination, non-proprietary nature, and license-free status (Tauberer, 2014). Bogdanovic-Dinic, Veljkovic, and Stoimenov (2014) introduced a model applying these principles to assess data openness in practical scenarios. Dias (2019) conducted a bibliometric study on e-government research, particularly in Ibero-American (IA) countries. In another study, Napitupulu's (2021) emphasized in less-explored areas on e-government research and identified future study avenues. Furthermore, Zhang, Hua, and Yuan (2017) revealed the global landscape of OGD research, emphasising the dominance of the UK and USA. Kulsum et al. (2022) explored OGD implementation in Indonesia and the Philippines, emphasising transparency, democracy, and citizen participation. Biswas and Chakraborty (2022) also showed that India too promotes OGD to manage and disseminate government data publicly. Through the study, they have illustrated the social value of OGD initiatives in India by demonstrating the availability of data and their implementations for the benefit of society. A comparative analysis of open government data initiatives in SAARC countries discussed by Biswas (2022) explored the transparency and openness of various activities of national governments of SAARC countries. On other hand, Catone (2023) employed network and factorial analysis to identify emerging trends in OGD studies and Erdayani, Afandi, and Afandi (2023) through their study on Open Government Partnership (OGP) bibliometric highlighted partnerships with multilateral institutions to promote open government.

The literature review indicates a rapid expansion in OGD research, with a focus on critical themes such as data quality, privacy, and data reuse. While foundational principles and assessment models have provided a solid framework, the global landscape, national implementations, and emerging trends are key focal points for researchers. However, the publications on OGD have not yet been assessed. This comprehensive review aims to fill gaps in the existing literature, offering valuable insights into the trajectory of OGD research over the past twelve years and paving the way for future investigations in this dynamic field.

3. Objectives

The primary aim of this research is to examine publications on OGD spanning from 2011 to 2022. The study focuses on the following key elements:

- Analysing the year-wise distribution of publications;

- Assessing the authorship pattern and Degree of Collaboration;
- Identifying the most prolific authors, core journals and most productive countries.
- Reveal the distribution of document types.
- Testing the verbal formulations of Bradford's law of scattering over the journal publications; and
- Examining the year-wise distribution of citations.

4.   Methodology

The present study is made to examine the trends in the growth of OGD literature by application of bibliometric indicators and other statistical tools. Bibliographic records for the period of 12 years starting from 2011 to 2022 were obtained from the Scopus database by using "open government data" in the "Title", "Abstract" and "Keyword" search criteria. The result was refined using the Language as "English" and the Publication Year as " > 2010 AND < 2023".

The search resulted in 1129 records and these were exported in the CSV format. Besides Ms-Excel, scientometric tool like BibExcel were used for the analysis of the records. Scopus was also used to find the h-index of the author and Cite Score 2021 of the journal.

5.   Analysis and results

Collected data have been tabulated, analysed and interpreted below.

5.1 *Year-wise distributions of publications*

Table 1 presents the yearly distribution of publications and the annual growth of OGD from 2011 to 2022. The Annual Growth Rate (AGR) is computed using the formula provided by Singh and Singh (2018):

$$AGR = \frac{Publications\,in\,final\,year - Publications\,in\,previous\,year}{Publications\,in\,previous\,year} \times 100$$

Table 1: Year-wise research output and growth

Year	No. of Publications	Cumulative total	%	AGR
2011	25	25	2.21	-
2012	34	59	3.01	36.00
2013	58	117	5.14	70.59
2014	75	192	6.64	29.31
2015	72	264	6.38	-4.00
2016	104	368	9.21	44.44
2017	124	492	10.98	19.23
2018	123	615	10.89	-0.81
2019	137	752	12.13	11.38
2020	126	878	11.16	-8.03
2021	121	999	10.72	-3.97
2022	130	1129	11.51	7.44
Total	1129	Average	94.08	

The table shows that over twelve years (2011-2022), there is a total of 1129 publications with an average of 94.08 publications per year. The highest number of publications occurred in 2019 with 137 (12.13%), while the lowest was in 2011 with 25 (2.21%). The growth rate indicates an overall steady increase, although fluctuations exist. Seven years experienced positive growth, while four had negative growth. The maximum growth rate was in 2013 at 70.59%, and the minimum was in 2020 at -8.03%.

5.2 *Authorship pattern and degree of collaboration*

Table 2 shows the trend of authorship pattern and the degree of collaboration (DC). The DC was calculated by dividing the number of multi-authored publications by the total number of publications (Subramanyam, 1983).

Table 2: Authorship pattern

Year	No Author	No. of SAP	% of SAP	No. of MAP	% of MAP	TP	DC
2011	4	5	20.00	16	64.00	25	0.64
2012	2	4	11.76	28	82.35	34	0.82
2013	4	8	13.79	46	79.31	58	0.79
2014	5	13	17.33	57	76.00	75	0.76
2015	6	10	13.89	56	77.78	72	0.78
2016	5	13	12.50	86	82.69	104	0.83
2017	7	16	12.90	101	81.45	124	0.81
2018	9	20	16.26	94	76.42	123	0.76
2019	5	16	11.68	116	84.67	137	0.85
2020	11	15	11.90	100	79.37	126	0.79
2021	5	13	10.74	103	85.12	121	0.85
2022	8	9	6.92	113	86.92	130	0.87
Total	71 (6.29%)	142	12.58	916	81.13	1129	

SAP = Single Author Publications, MAP = Multi- Author Publications, TP = Total Publications

Table 2 shows that out of 1129 publications, 71 publications (6.29%) have no author name, 142 (12.58%) are single-authored and 916 (81.13%) are multi-authored papers. The DC varies from 0.64 to 0.87 which indicates a collaborative research trend in OGD research.

5.3 *Most prolific authors*

Table 3 shows the top 10 prolific authors with their institutes, h-index, total publications on this particular topic, citations received by those publications, citations per publication (CPP), and rank according to CPP. CPP is computed as the average number of citations per publication (Li & Ho, 2008). H-index of different authors was collected from the author profile of Scopus (Scopus, n.d.).

Table 3: Most prolific authors

Name	Affiliation	h index	TP	TC	CPP	Rank
Janssen, M.	Delft University of Technology, Netherlands	59	33	645	19.55	3
Zuiderwijk, A.	Faculteit Techniek, Bestuuren Management, Netherlands	26	32	568	17.75	4
Saxena, S.	Graphic Era Deemed to be University, India	12	20	141	7.05	8
Alexopoulos, C.	University of the Aegean, Greece	11	13	202	15.54	6
Nikiforova, A.	Tartu Ülikool, Estonia	9	12	107	8.92	7
Viterbo, J.	Universidade Federal Fluminense, Brazil	11	12	64	5.33	9
Charalabidis, Y.	University of the Aegean, Greece	26	11	179	16.27	5
Crusoe, J.	Göteborgs Universitet, Sweden	4	11	38	3.45	10
Kalampokis, E.	University of Macedonia, Greece	13	11	276	25.09	1
Tarabanis, K.	University of Macedonia, Greece	28	11	276	25.09	1

TP = Total Publications, TC = Total Citations, CPP = Citations per publication

It is apparent from table 3 that Janssen, M published most publications on this topic and he also received the maximum citations among the top 10 prolific authors. His h-index, 59 is also the highest among all. In terms of CPP, Kalampokis, E and Tarabanis, K jointly ranked the top. Among the top 10 prolific authors, 4 are from Greece.

5.4 *Document types*

Table 4: Distribution of document types

Document type	TP	%
Conference Paper	549	48.63
Article	431	38.18
Conference Review	70	6.20
Book Chapter	54	4.78
Review	18	1.59
Note	4	0.35
Book	1	0.09
Data Paper	1	0.09
Erratum	1	0.09

Table 4 illustrates the distribution of document types in open government data research. The findings indicate a diverse use of formats, with conference papers, articles, conference reviews, book chapters, and reviews collectively comprising 99.38% of the total. Conference papers lead with 48.63% preference, closely followed by articles at 39.18%. In contrast, data papers and errata have minimal popularity, each accounting for only 0.09% of the overall document landscape.

5.5 *Distribution of core journals*

Table 5 reflects the top 10 core journals with their Cite Score 2021 (Scopus, n.d.-a) and CPP. A modified rank of the top 10 popular journals was also made based on the CPP.

Table 5: Distribution of core journals

Name	Publisher	Cite Score 2021	TP	TC	CPP	Rank
Government Information Quarterly	Elsevier	14.5	45	2756	61.24	1
eJournal of eDemocracy and Open Government (JeDEM)	Department for E-Governance and Administration, Danube University Krems, Austria	1.1	25	249	9.96	7
Transforming Government: People, Process and Policy	Emerald Publishing	4.2	18	363	20.17	4
Information Polity	IOS Press	4.4	17	420	24.71	2
Sustainability	Multidisciplinary Digital Publishing Institute (MDPI), Switzerland	5	13	59	4.54	10
International Journal of Electronic Government Research	IGI Global Publishing	2.4	9	56	6.22	8
Public Performance & Management Review	Taylor & Francis	3.9	7	147	21.00	3
Big Data and Society	SAGE	8.1	5	78	15.60	5
Digital Policy, Regulation and Governance	Emerald Publishing	3.7	6	64	10.67	6
Electronic Library	Emerald Publishing	3.1	5	25	5.00	9

The table highlights the productivity of journals, with 'Government Information Quarterly' leading, followed by 'eJournal of eDemocracy and Open Government (JeDEM)' and others. 'Government Information Quarterly' boasts the highest CiteScore in 2021 (14.5), while the 'International Journal of Electronic Government Research' has the lowest (2.4). 'Government Information Quarterly' received the maximum citations (2756), whereas 'Electronic Library' received the lowest (25). In terms of CPP, 'Government Information Quarterly' retains the 1st position, but 'Information Polity' secures the 2nd position, and 'Public Performance & Management Review' takes the 3rd position.

5.6 *Applications of Bradford's Law of Scattering*

A total of 200 journals contributed 431 publications which received 8268 citations during the study period. For testing the verbal formulations of Bradford's Law of Scattering, 200 journals are divided equally into 3 zones. Table 6 shows the distribution of journals and the corresponding number of citations in the three zones along with the value of Bradford multiplier.

Table 6: Scattering of journals and citations over Bradford zones

Zones	Journals	Citations	Bradford Multiplier
I	1	2744	
II	19	2684	19
III	180	2840	9.47
Total	200	8268	14.24

The identified zones are arranged in the geometric series as 1:n:n2, where n is the value of the Bradford multiplier (Savanur, 2019). In this series only 1 journal was found in the nucleus and the mean Bradford Multiplier is 14.24. So, 1:n:n2 = 1:(1*14.24):(1*14.242) = 1:14.24:202.78. According to this calculation, number of total publications is 218.02. The Error percentage is 9.01%. Keeping a 5% degree of significance it can be concluded that Bradford's Law of Scattering does not conform to observed data of scattering of open government data research in different journals during 2011-2022.

5.7 *Most productive countries*

Table 7 shows the top 10 productive countries with the number of total publications, total citations, and the value of CPP and RCI. Relative Citation Impact (RCI) is computed by dividing a country's share of total citations by a country's share of total publications to measure both the influence as well as the visibility of research (Bhardwaj, 2017).

Table 7: Most productive countries

Country	TP	TC	CPP	RCI
United States of America	132	1516	11.48	1.08
Brazil	87	480	5.52	0.52
Netherlands	78	1069	13.71	1.29
China	75	375	5.00	0.47
United Kingdom	69	1268	18.38	1.73
India	60	256	4.27	0.40
Germany	56	1015	18.13	1.70
Italy	48	498	10.38	0.97
Taiwan	47	298	6.34	0.60
Greece	40	496	12.40	1.16
Total	1129	12022	10.65	1.00

The table reflects that the USA produces the maximum number of publications (132) and receivesthe maximum number of citations (1516). The value of CPP is highest for the UK as 69 publications of this country received 1268 citations during this period. The value of CPP is lowest for India as 60 publications in this country received only 256

citations. The value of RCI reveals that among the top 10 productive countries, the high impact of research is visible in 5 countries. The UK shows the highest impact of research and India shows the lowest impact of research.

5.8 *Year-wise distribution of citations*

Table 8 displays the distribution of citations by year. Among the 1129 publications, 307 (27.19%) received no citations, while 822 (72.81%) received at least one. The majority (274) garnered between 2 to 5 citations, and only 17 publications received over 100 citations. Over the study period, these publications accumulated 12,022 citations, resulting in an average of 10.65 CPP. Notably, publications from 2017 received the most citations, while those from 2022 received the fewest. The CPP was highest in 2011 and lowest in 2022.

Table 8: Year-wise distribution of citations

Year	0	%	1	2-5	6-10	11-20	21-50	51-100	>100	Total cited papers	%	TP	TC	CPP
2011	8	32.00	1	2	4	2	4	1	3	17	68.00	25	688	27.52
2012	6	17.65	3	6	6	2	7	3	1	28	82.35	34	716	21.06
2013	10	17.24	4	9	10	11	11	2	1	48	82.76	58	895	15.43
2014	14	18.67	6	15	11	9	14	5	1	61	81.33	75	1207	16.09
2015	15	20.83	8	21	8	7	6	4	3	57	79.17	72	1512	21.00
2016	21	20.19	8	29	16	11	9	6	4	83	79.81	104	1809	17.39
2017	20	16.13	13	34	17	12	18	6	4	104	83.87	124	1909	15.40
2018	25	20.33	16	35	25	11	9	2	0	98	79.67	123	900	7.32
2019	28	20.44	13	46	26	15	8	1	0	109	79.56	137	903	6.59
2020	30	23.81	17	39	18	11	9	2	0	96	76.19	126	928	7.37
2021	46	38.02	25	22	16	9	3	0	0	75	61.98	121	429	3.55
2022	84	64.62	24	16	5	1	0	0	0	46	35.38	130	126	0.97
Total	307	27.19	138	274	162	101	98	32	17	822	72.81	1129	12022	10.65
%	27.19		12.22	24.27	14.35	8.95	8.68	2.83	1.51					

6.   Discussion

The critical analysis of the presented data highlights the dynamic and evolving nature of open government data research. The field has experienced steady growth, but the fluctuations in growth rates, authorship patterns, journal preferences, and deviations from theoretical models like Bradford's Law warrant careful consideration. These variations can guide future research directions, foster collaboration, and enhance the impact of open government data research on evidence-based decision-making and transparent governance. Further investigations into the reasons behind negative growth years, authorship dynamics, and the impact of publications from different countries can contribute to the refinement of strategies and methodologies in this interdisciplinary field.

7.   Conclusion

From 2011 to 2022, research outputs on open government data exhibited consistent growth, resulting in a total of 1129 publications. The initial count in 2011 was

only 25 publications, which increased to 130 by 2022. Negative growth rates were observed in 2015, 2018, 2020, and 2021. Collaboration was evident in multi-authored publications, and the USA led in research output with 132 publications. However, the UK topped in terms of CPP and RCI. Most publications received at least one citation during the study period, with only 27.19% receiving none. Conference papers were the preferred document type, and the dataset did not conform to Bradford's Law of Scattering.

The future of open government data research is very promising. This advancement will help with evidence-based decision-making, transparency, and the efficient functioning of governments. The steady growth in research publications over recent years suggests sustained interest and activity on this particular topic. This work can provide valuable insights into research trends and help identify key themes and topics in the field.

References

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly,* 32(4), 399-418. Retrieved from https://doi.org/10.1016/j.giq.2015.07.006

Biswas, A. (2022). A comparative analysis of open government data initiatives in SAARC countries. In Ashis Biswas & Mayuri Das Biswas (Eds.), *Panorama of open access: progress, practices and prospects* (pp. 158 - 177). New Delhi: Ess Ess Publications. Retrieved from https://www.researchgate.net/publication/359511471_A_Comparative_Analysis_of_Open_Government_Data_Initiatives_in_SAARC_Countries

Biswas, A., & Chakraborty, C. (2022). Towards openness: a study on open government data initiative in India. *College Libraries,* 37(1), 22-33. Retrieved from http://college libraries.in/index.php/CL/article/view/63

Bhardwaj, R.K. (2017). Mapping of global research literature in the field of cloud computing. *Collnet Journal of Scientometrics and Information Management,* 11(2), 173-91.Retrieved from http://doi.org/10.1080/09737766.2017.1284727

Bogdanovic-Dinic, S., Veljkovic, N., & Stoimenov, L. (2014). How open are public government data an assessment of seven open data portals. In M. Rodriguez-Bolivar (Ed.), *Public administration and information technology* (pp. 25-44). Retrieved from https://doi.org/10.1007/978-1-4614-9982-4_3

Broadus, R. N. (1987). Toward a definition of "bibliometrics." *Scientometrics,* 12(5-6), 373-379. Retrieved fromhttps://doi.org/10.1007/bf02016680

Catone, M. C. (2023). The role of open data in digital society: the analysis of scientific trending topics through a bibliometric approach. *Frontiers in Sociology,* 8, 1-10. Retrieved from https://doi.org/10.3389/fsoc.2023.1134518

Data.gov Home.(n.d.). Retrieved from https://data.gov/

Dias, G. P. (2019). Fifteen years of e-government research in Ibero-America: a bibliometric analysis. *Government Information Quarterly,* 36(3), 400-411.Retrieved from https://doi.org/10.1016/j.giq.2019.05.008

Erdayani, R., Afandi, M., & Afandi, S. A. (2023). Bibliometric analysis of open government: a study on the open government partnership. *Indo-Fintech Intellectuals: Journal of Economics and Business,* x 3 (2), 276-294. Retrieved from http://doi.org/10.54373/ifijeb.v3i2.180

Find open data - data.gov.uk. (n.d.). Retrieved from https://www.data.gov.uk/

Home.(n.d.). Retrieved from https://data.gov.in/

Junior, L. A. F., De Almeida Guimaraes, L. G., Segundo, E. M. D., Da Costa, W. P. L. B., & El-Aouar, W. A. (2020). Open government and e-government: concepts, gaps,

interfaces, and trends. In A. Farazmand (Ed.), *Springer eBooks,* 1-11. Retrieved fromhttps://doi.org/10.1007/978-3-319-31816-5_4098-1

Kulsum, U., Nurmandi, A., Muallidin, I., Loilatu, M. J., &Kurniawan, D. (2022). A bibliometric analysis of open government: study on Indonesia and Philippines. *Journal of Governance,* 7(1), 133-143. Retrieved from https://doi.org/10.31506/jog.v7i1.13934

Li, Z., & Ho, Y-S. (2008). Use of citation per publication as an indicator to evaluate contingent valuation research. *Scientometrics,* 75(1), 97-110. Retrieved fromhttp://doi.org/10.1007/s11192-007-1838-1

Napitupulu, D. (2021). A bibliometric analysis of e-government.Research. *Library Philosophy and Practice* (e-journal). Retrieved from https://digitalcommons.unl.edu/libphilprac/5861

National Data Sharing and Accessibility Policy (NDSAP) - 2012. [pdf]. (2012). Retrieved from https://data.gov.in/sites/default/files/NDSAP.pdf

Open Government Data - OECD. (n.d.). Retrieved from https://www.oecd.org/gov/digital-government/open-government-data.htm

Tauberer, J. (2014). Open government data definition: the 8 principles of open government data. In *Open Government Data: The Book* (2nd ed.). Retrieved fromhttps://opengovdata.io/2014/8-principles/

Savanur, K. P. (2019). Application of Bradford's law of scattering to the economics literature of India and China: a comparative study. *Asian Journal of Information Science and Technology,* 9(1),1-7. Retrieved from http://dx.doi.org/10.51983/ajist-2019.9.1.2610

Scopus. (n.d.). Retrieved from https://www.scopus.com/freelookup/form/author.uri Scopus. (n.d.-a). Retrieved from https://www.scopus.com/sources.uri?zone=Top NavBar&origin=searchauthorfreelookup

Singh, P.K., & Singh, A.P. (2018).Growth trend in global big data research publications as seen from SCOPUS database. *Professional Journal of Library and Information Technology,* 8, 49-61. Retrieved from https://www.researchgate.net/publication/332093972_Growth_ Trend_in_Global_ Big_Data_Research_Publications_as_Seen _from_SCOPUS_Database

Subramanyam, K. (1983). Bibliometric studies of research collaboration: a review. *Journal of Information Science,* 6(1), 33-38. Retrieved from http://dx.doi.org/10.1177/0165551 58300600105

The Rio declaration on environment and development. [pdf]. (1992). Retrieved from https://www.iau-hesd.net/sites/default/files/documents/rio_e.pdf

Zhang, Y., Hua, W., & Yuan, S. (2017). Mapping the scientific research on open data: a bibliometric review. *Learned Publishing,* 31(2), 95-106. Retrieved from https://doi.org/10.1002/leap.1110