IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL)
ISSN (P): 2347–4564; ISSN (E): 2321–8878
Vol. 10, Issue 3, Mar 2022, 33–42

GROWTH OF BIG DATA RESEARCH PRODUCTIVITY: A SCIENTOMETRIC ANALYSIS

Ravindranath Wodeyar[1] & Dr. K R Mulla[2]

[1]Librarian, Proudhadevaraya Institute of Technology, Hosapete

[2]Librarian, Visveswaraya Technological University. Belagavi, Karnataka

Received: 02 Mar 2022 **Accepted:** 09 Mar 2022 **Published:** 10 Mar 2022

ABSTRACT

This study examined the growth rate of Big Data research literature over the period 2001 to 2020. Data were extracted from WoS and Scopus Databases and merged with Bibliometrics, R programming. Collected data further refined and remove duplicate records and finally analyzed a total of 19667 research papers. This study aims to determine various scientometric indicators, including the year-wise distribution of records, annual growth rate, compound annual growth rate, authorship pattern, etc., This article shows an increase in publications from 0.005 to 21.37% with an annual growth rate of 89.53% and a CAGR of 41.56%. Over the study period, the results reported here confirm that the relative growth rate decreased and the doubling time increased. Writing modeling showed that 93.66% of articles were co-authored. As the results show, the growth rate of big data research is at an alarming rate.

KEYWORDS: *Scientometrics; Big Data, Bibliometrix, Authorship Pattern*

INTRODUCTION

Big data is considered a buzzword in business and industry (Vossen, 2014). "Big data" originally referred to managing, handling, and analyzing very large datasets and has been used to refer to this ever since the mid-1990s. The term 'Big data' was coined in 1990 by John Mashey, (Diebold, 2012). In the age of the World Wide Web and Web 2.0 technologies, a constant amount of structured and unstructured data is generated from various sources, including email, social media platforms such as Facebook, WhatsApp, LinkedIn, blogs, online transactions, articles, and forums. In addition, different types of sensor data are generated from different sources such as health sciences, environmental organizations, metrological departments, business data, census data, company data, etc. in larger volume, and enormous velocity is called Big data. The scientometric technique is a widely recognized quantitative tool for identifying and measuring the publication growth in any subject. Scientometrics is a quantitative discipline in which a large number of studies are conducted on numerical analysis of many aspects of the literature on a particular topic. It statistically analyses published content using aspects of bibliographic data. In recent decades, scientometric studies have received much attention and are widely used to evaluate scientists' research and the growth of many science disciplines (Verma and Shukla, 2020).Scientometrics can also be used to identify new areas of research. **Accordingly,** the **present** study was **performed** to determine the growth of **the** literature in Big Data, **the** annual growth rate, **the** compound annual growth rate, and **the** collaborative research.

METHODOLOGY

The purpose of this study is to use scientometric indicators to examine the growth of literature on the topic of "Big Data."

- **Databases**: To obtain the data for the specified aims, WoS and Scopus databases were exploited.
- **Period**: 2001 to 2020, twenty years.
- **Search string**: "Big Data" in topic field and limited to articles, conference papers, and review reports.
- **Sample size**: The research **examined** a total of 19667 records.
- **Analysis and Visualization Tool**: The downloaded data was also saved in BibTeX files, which were then imported, merged, and removed from duplicatesin Bibliometrix: R Programming and the results were tabulated.

OBJECTIVES

- To analyzes the Annual Growth Rate of Publications
- To determine the Relative Growth rate and doubling time
- To find out Compound Annual Growth Rate
- To assess the Authorship Pattern

LITERATURE REVIEW

(Jin et al., 2015) remarks 'Big Data has become an increasingly popular term, and it refers to a very significant area of research. When compared to traditional data, the attributes of 'Big Data' are described by 5V that represents for huge Volume, high Velocity, high Variety, low veracity, and high Value

(Manyika et al., 2011) focused on, today large and complex datasets are collected through a variety of channels for many reasons. The technology used to generate this huge data includes mobile devices, remote sensing, software logs, wireless sensor networks, social media, and so on. Scientists and businessmen required new theories, methods, and analytics tools to deal with the 'Big Data'

Many research studies have been conducted on scientometric mappings of research activities for a particular area of research. However, previous works on diverse disciplines and narrow topics helped us in formulating our research plan. (Singh et al., 2015)Previous studies identified three main areas of scientometric mapping work: (a) scientometric mapping of a particular field of study with or without a specific focus on a particular country/region (b) scientometric mapping of research in a narrow field with or without a focus on a particular country/ region; (c) a comparative study of a research organization or country in a specific subject area.

(Inamdar et al., 2020) emphasizes systematic literature review and bibliometric analysis of Big Data Analysis adoption in the supply chain and its applications in diverse industries from 2014 to 2018. Several countries and sectors have been examined in this paper by BDAA studies. Furthermore, the paper examines different tools and techniques used in BDAA studies.

(Pinarbasi & Canbolat, 2020) looks at the bibliometrics of publications on big data in indexed marketing journals tooto examine how the concept of big data is evaluated in marketing literature. In this study, descriptive statistics are first presented, followed by the top-ranked journals, authors, and countries that contribute to each of the authors. In addition, the study identifies the most influential studies for big data concept-setting literature. (Kalantari et al., 2017) draws attention to the past few years, the explosive growth of mobile, social media, the Internet of Things, and other data source haves led to big data's emergence. Specifically, this paper examines the worldwide research trends concerning big data and the most relevant areas within it.

DATA ANALYSIS AND INTERPRETATION

Research Output in Big Data Research

Figure 1 illustrates study results annually from 2001 to 2020. Publication output in Big Data research increased from 1 (0.005%) in 2001 to 4 200 (21.37%) in 2021. It can be seen that in the first decade (2001-2010) only a limited number of publications were published and in the second decade (2011-2020), this trend is increasing.

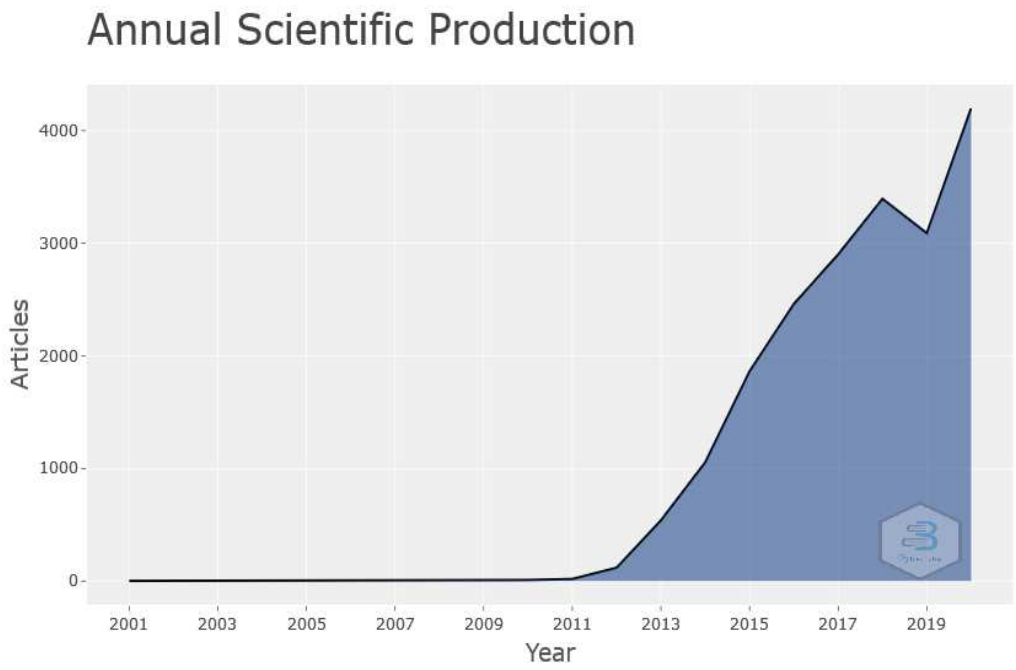

Figure 1

Annual Ratio of Growth (ARoG)

The annual growth and distribution patterns of publications for the period 2001 to 2020 are given in Table 1.Theannual ratio of growth is calculated with the publications of the current year divided by the publications of the previous year.

From Table 1 it can be seen that in 2001, the total number of publications in Big Data was 1. In 2020, this number increased to 19667. During this period the annual ratio of growth ranges between 0.91 and 1.09. The annual ratio of growth thus calculated shows that there is steady growth over the past five years. It varies between 0.93 and 1.03.

Table 1: World research output Annual Ratio of Growth of Big Data

Year	Number of Publications	Percentage	Cumulative Papers	Cumulative Percentage	ARoG
2001	1	0.005	1	0.005	
2002	2	0.010	3	0.015	2
2003	1	0.005	4	0.02	0.5
2004	1	0.005	5	0.025	1
2005	1	0.005	6	0.03	1
2006	2	0.010	8	0.01	2
2007	1	0.005	9	0.005	0.5
2008	4	0.020	13	0.067	4
2009	8	0.041	21	0.107	2
2010	9	0.046	30	0.152	1.13
2011	18	0.092	48	0.244	2
2012	118	0.600	166	0.844	6.56
2013	538	2.736	704	3.58	4.56
2014	1054	5.359	1758	8.94	1.96
2015	1862	9.468	3620	18.41	1.77
2016	2461	12.513	6081	30.92	1.33
2017	2901	14.751	8982	45.67	1.18
2018	3395	17.262	12377	62.93	1.18
2019	3090	15.712	15467	78.64	0.918
2020	4200	21.356	19667	100	1.36

Relative Growth Rate (RGR) and Doubling Time (Dt)

RGR means the increase in the number of publications per unit of time. It is also called the continuous growth rate concerning scientific literature publication. The growth rate of all publications as has been measured based on RGR and Dt model, which was developed by Mahapatra in 1985. (Mahapatra, 1985) The formula used to calculate the relative growth rate and doubling time is:

RGR = (ln W2 - ln W1)/(t2 - t1)

Where **RGR** means the relative growth rate over a specified period of interval

ln W1=Log wl= Natural log of the initial number of publications

ln W2=Log W2 =Natural he of the final number of publications

T1 = The unit of initial time

T2 = The unit of the final time

Doubling Time

The doubling time is the given period required for a quantity to double in size or value. It is directly related to RGR, where RGR is constant. The quantity undergoes exponential growth and has a constant doubling time or period which can be calculated directly from the growth rate. So the Doubling time is calculated by using the Formula; **Dt=0.693/R**

Where,

Dt = Doubling Time

R= Growth rate

Table 2 depicts the relative growth rate and doubling time of Big Data publishing from 2001 to 2020. The RGR was lowest in 2007 with0.118and highest in 2013 with 1.445. The mean relative growth rate in the first four years (2001 to 2004) was 0.536. Over the next four years (2005 to 2008) the growth rate was decreased slightly to 0.239. Furthermore, it rebounded to 0.from7in 2009 to 2012. In 2013-2016 it was again increased to 0.9 and last quarter it is decreased to 0.293.

The doubling time shows oscillation and peaks in 2007 with5.884. The mean doubling time in the first four years (2001 to 2004) was 2.048 and it was increased to the highest doubling time of 3.495 in the second year (2005 to 208. From there can be seen that the relative growth rate has decreased and the doubling time has increased.

Table 2: Relative Growth Rate and Doubling Time

Year	Number of Publications	Cumulative Publications	Cumulative Percentage	W1	W2	RGR	Mean RGR	DT	Mean DT
2001	1	1	0.01	-	0	-			
2002	2	3	0.02	0	1.099	1.099	0.536	0.631	2.048
2003	1	4	0.02	1.099	1.386	0.288		2.409	
2004	1	5	0.03	1.386	1.609	0.223		3.106	
2005	1	6	0.03	1.609	1.792	0.182	0.239	3.801	3.495
2006	2	8	0.04	1.792	2.079	0.288		2.409	
2007	1	9	0.05	2.079	2.197	0.118		5.884	
2008	4	13	0.07	2.197	2.565	0.368		1.885	
2009	8	21	0.11	2.565	3.045	0.48	0.637	1.445	1.355
2010	9	30	0.15	3.045	3.401	0.357		1.943	
2011	18	48	0.24	3.401	3.871	0.47		1.474	
2012	118	166	0.84	3.871	5.112	1.241		0.559	
2013	538	704	3.58	5.112	6.557	1.445	0.9	0.48	0.883
2014	1054	1758	8.94	6.557	7.472	0.915		0.757	
2015	1862	3620	18.41	7.472	8.194	0.722		0.959	
2016	2461	6081	30.92	8.194	8.713	0.519		1.336	
2017	2901	8982	45.67	8.713	9.103	0.39	0.293	1.777	2.483
2018	3395	12377	62.93	9.103	9.424	0.321		2.161	
2019	3090	15467	78.64	9.424	9.646	0.223		3.109	
2020	4200	19667	100.00	9.646	9.887	0.24		2.885	

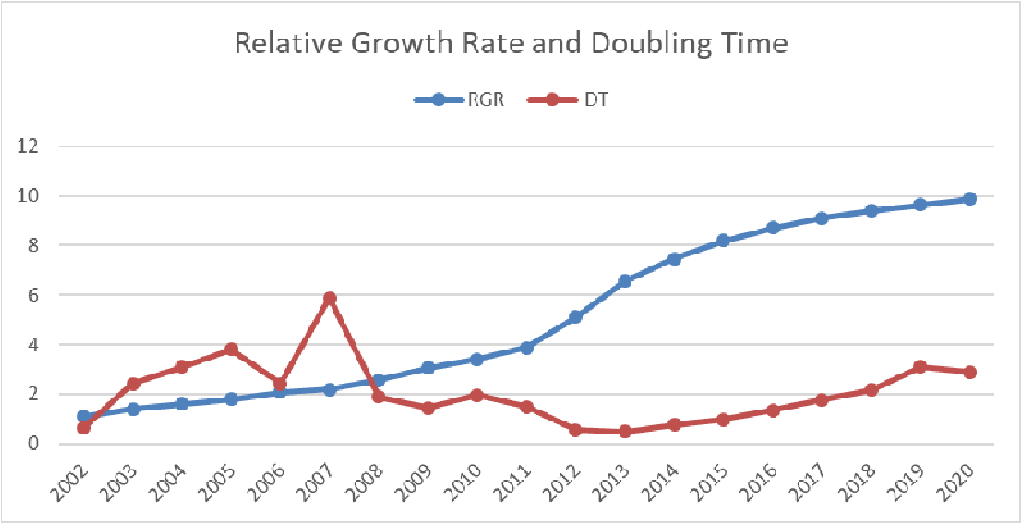

Figure 2: Relative Growth Rate and Doubling Time.

The Annual Growth Rate of Publications

Table 3 shows the annual growth rate (AGR) of research output for the specified research period, during which the maximum annual growth rate is determined in 2012, i.e. 55.556, followed by 355.932 in 2013. Further, the table shows an average annual growth rate of 89.529. The annual growth rate is calculated according to the formula suggested by Kumar and Kaliyaperumal, 2015 and mentioned below:

$$AGR = \frac{\text{End Value-First Value}}{\text{First Value}} \times 100$$

Table 3: Annual Growth Rate of Publications

Year	Number of Publications	Percentage	AGR
2001	1	0.005	000.000
2002	2	0.010	100.000
2003	1	0.005	-050.000
2004	1	0.005	000.000
2005	1	0.005	000.000
2006	2	0.010	100.000
2007	1	0.005	-050.000
2008	4	0.020	300.000
2009	8	0.041	100.000
2010	9	0.046	012.500
2011	18	0.092	100.000
2012	118	0.600	555.556
2013	538	2.736	355.932
2014	1054	5.359	095.911
2015	1862	9.468	076.660
2016	2461	12.513	032.170
2017	2901	14.751	017.879
2018	3395	17.262	017.029
2019	3090	15.712	-008.984
2020	4200	21.356	035.922
	19667	100	**89.529**

The Ratio of Growth and Compound Annual Growth Rate of Publications

Table 4 describes the compound annual growth rate of Big Data publications over a period. The compound annual growth rate is measured by taking the n^{th} root of the total percentage growth rate, where n is the number of years in the period (Subramanyam, 1983). It can be seen that the CAGR was recorded in the year 2002 with 100, followed by 71.23 in 2015. The table also shows a compound annual growth rate of 41.551.

The compound annual growth rate **is** calculated **according to** the following formula (Shukla, 2020).

$$CAGR = [(\text{Ending Value} / \text{Beginning Value})1/n - 1]$$

Table 4: Ratio of Growth and Compound Annual Growth Rate of Publications

Year	Number of Publications	Percentage	CAGR
2001	1	0.005	
2002	2	0.01	100
2003	1	0.005	0
2004	1	0.005	0
2005	1	0.005	0
2006	2	0.01	14.87
2007	1	0.005	0
2008	4	0.02	21.9
2009	8	0.041	20.68
2010	9	0.046	27.65
2011	18	0.092	33.51
2012	118	0.6	54.3
2013	538	2.736	68.87
2014	1054	5.359	70.82
2015	1862	9.468	71.23
2016	2461	12.513	68.3
2017	2901	14.751	64.59
2018	3395	17.262	61.32
2019	3090	15.712	56.27
2020	4200	21.356	55.13
Total	19667	100	41.551

COLLABORATIVE RESEARCH

Collaboration allows individuals to research together to achieve a specified and regular purpose (Dillenbourg, 1999). Table 5 shows the majority (86.56%) of publications published by multi-authorship. It is seen that 13.95% of the publications are made by a single author, 21.76% of publications by two authors, 20.80% of contributions by three authors, and 16.88% of publications were contributed by four authors. In addition, 63.86% of publications are written by more than four authors. The most forms of collaboration were ten or more authors (20.08%), six authors (10.7%), four authors (10.86%), and five authors (10.92%) respectively. Therefore there is a tendency to collaborate in research.

Table 5: Authorship Pattern

Authorship	Frequency of Publications	Percentage	Cumulative Frequency of Publications	Percentage
Single Author	2744	13.952	2744	13.952
Two Authors	4514	22.952	7258	36.904
Three Authors	4319	21.961	11577	58.865
Four Authors	3321	16.886	14898	75.751
Five Authors	2093	10.642	16991	86.393
Six Authors	1266	6.437	18257	92.831
Seven Authors	594	3.020	18851	95.851
Eight Authors	265	1.347	19116	97.198
Nine Authors	187	0.951	19303	98.149
Ten Authors	90	0.458	19393	98.607
>Ten Authors	274	1.393	19667	100.000
Total	19667	**100**		

FINDINGS AND CONCLUSION

Scientometric studies have developed a body of theoretical knowledge and a group of techniques and applications based on the distribution of bibliographic data. The widespread adoption of Scientometric techniques led to the development of a new and more precise technique. It is hoped that the ongoing theoretical work will pave the way for more innovative techniques. The study examines the growth of publications, annual growth rate, compound annual growth rate, authorship pattern in Big Data literature. The growth of publications was ranged from 0.005in the year 2001 to 21.37 in the year 2020. From the year2001 to 2010, we can found a very slow growth of publications productivity. The study found that there is an increasing trend during the second decade i.e 2011 to 2020.

The overall annual growth rate was 89.93 during the study. The highest annual growth rate was observed in 2012 at 555.55%. The relative growth rate was decreasing and the doubling time was increasing from 2001 to 2020. The compound annual growth rate was 41.55. The authorship pattern shows that 22.95% of the publications were contributed by more than two authors and this result shows the collaborative network is high in Big Data literature.

REFERENCES

1. *Aria M, Cuccurullo C (2017). "bibliometrix: An R-tool for comprehensive science mapping analysis." Journal of Informetrics, 11(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007.*
2. *Arun Kumara, T.S. & Santhosh Kumar,K.T. (2020) Scientometric analysis of literature on Gravity. Library Philosophy and Practice (e-journal), 4024.*
3. *Arya, Chanda& Sharma, Superma.(2011). Authorship trends and collaborative research in veterinary sciences: A bibliometric study. Chinese Lib.: An Int. Electr. J.34. http://www.iclc.us/cliej/cl34AS.pdf/.*
4. *Brij Mohan Gupta, et al. (2013)World cataract research: A scientometric analysis of publications output during 2002-11. Lib. Philo. Prac. (e-journal), paper 895.*
5. *Diebold, Francis X., On the Origin(s) and Development of the Term 'Big Data' (September 21, 2012). https://ssrn.com/abstract=2152421 or http://dx.doi.org/10.2139/ssrn.2152421*
6. *Dillenbourg, P. (1999). What do you mean by collaborative learning?" collaborative- learning: cognitive and computational approaches. Elsevier, 1–19*
7. *Dutta, B. &Rath, D.S. (2013). Cosmology research in India: A scientometric study. Lib. Philo. Pract., (e-journal), paper 996.*
8. *Falagas, Matthew E.; Papastamataki, Paraskevi A. &Bliziotis, Ioannis A. (2006) A bibliometric analysis of research productivity in parasitology by different world regionsduring a 9-year period (1995-2003). BMC Infectious Diseases, 6(56). http://www.biomedcentral.com/1471-2334/6/56/.*
9. *Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2020). A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. In Journal of Enterprise Information Management (Vol. 34, Issue 1). https://doi.org/10.1108/JEIM-09-2019-0267*

10. *Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. Big Data Research, 2(2), 59–64. https://doi.org/10.1016/J.BDR.2015.01.006*
11. *Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ale Ebrahim, N., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017). A bibliometric approach to tracking big data research trends. Journal of Big Data, 4(1). https://doi.org/10.1186/S40537-017-0088-1*
12. *Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung A. (2011). Big data: The next frontier for innovation, competition, and productivity. Technical Report, McKinsey Global Institute. www.mckinsey.com/mgi.*
13. *Pinarbasi, F., & Canbolat, Z. N. (2020). Business Ecosystem & Strategy Big data in marketing literature : A bibliometric analysis. International Journal of Business Ecosysten & Strategy, 1(2).*
14. *Singh, V. K., Banshal, S. K., Singhal, K., & Uddin, A. (2015). Scientometric mapping of research on 'Big Data.' Scientometrics, 105(2), 727–741. https://doi.org/10.1007/s11192-015-1729-9*
15. *Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. Journal of information Science, 6(1), 33-38.*
16. *Van Raan, A.F.J. (1997). Scientometrics: State-of-the-art. Scientometrics, 38, 205-15.*
17. *Verma, M. K., & Shukla, R. (2020). Mapping the research trends on information literacy of selected countries during 2008-2017: A scientometric analysis. DESIDOC Journal of Library & Information Technology, 39(3), 125-130.*
18. *VossenGottfried (2014).Big data as the new enabler in business and other intelligence. Vietnam Journal of Computer Science volume 1, pages 3–14 (2014)*
19. *Web of Science. https://clarivate.com/webofsciencegroup/solutions/web-of-science/*