

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

2021

## Mapping of Research Output in the field of Big Data Analytics: Scientometric Study

Prachi Srivastava

*Central University of Punjab, prachiacademic@gmail.com*

Rupinder Singh

*Centre for Studies in Social Sciences, Calcutta, rupinder@outlook.in*

Sharad Kumar Sonkar

*Babasaheb Bhimrao Ambedkar University Lucknow, sksonker@yahoo.co.in*

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Scholarly Publishing Commons](#)

---

Srivastava, Prachi; Singh, Rupinder; and Sonkar, Sharad Kumar, "Mapping of Research Output in the field of Big Data Analytics: Scientometric Study" (2021). *Library Philosophy and Practice (e-journal)*. 6378.  
<https://digitalcommons.unl.edu/libphilprac/6378>

## Mapping of research output in the field of Big Data Analytics: Scientometrics study

Dr. Prachi Srivastava<sup>1</sup>

Mr. Rupinder Singh<sup>2</sup>

Dr. Sharad Kumar Sonkar<sup>3</sup>

1. Central University of Punjab, Ghudda Bathinda-151401
2. Centre for Studies in Social Sciences, Calcutta Kolkata, West Bengal- 700 094
3. Associate Professor Department of Library and Information Science Babasaheb Bhimrao Ambedkar University, Lucknow-226025

### Abstract:

This study aims to determine the scientometrics analysis of recent research output in the emerging field of Big Data Analytics from 2011 to 2020. The study of Big Data Analytics began in the last few years and has gained tremendous traction in a short period of time. It is now widely regarded as one of the most important developing study areas in computer sciences and related disciplines. The data has been obtained from the Scopus Database and excel and VOSviewer software used for the analysis; a total of 2304 research papers were evaluated. The study examines total output, growth in output, authorship and collaboration patterns at the country level, major contributors (countries, institutions, and individuals), top publication sources, analytical trends, and highly cited papers. The study explored continuous growth in the literature of Big data analytics.

**Keywords:** Research trends, Big data, Big data Analytics

**Introduction:** As our society progresses through this era of information and technology, our consumer-like behavior has made every commodity available to us as consumable goods. Just like minerals and rare metals, one another universally sought-after commodity is Data. Before we start exploring big data, we must start within ourselves. We will find that we are a part of the Big Data Ocean. We are generating data every second. When we think, we read, we write, when we shop online, when we use social media. Even when we use our personal assistants like Alexa, or Google home assistance to do the tiniest tasks for us, we are actually producing data.(Khan et al., 2014). "Data is the new oil." This term was coined by *Clive Humby* in ANA Senior marketer's summit, at Kellogg School in 2006(Palmer, 2006). This metaphor has clearly grown in popularity in recent years, but is it a sound way of looking at data? Is there a better approach to think about data that can help us better comprehend its function in the twenty-first century, particularly with advances in predictive analytics and artificial intelligence? Like oil, data is only valuable if it is in a usable form. Just like crude oil needs to be processed in refineries and turned into petroleum, raw data also needs to be refined, curated and processed

before it can be used for analytics or, or it can be put to any application (Mavuduru, 2020). But it is clear at the rate we are producing and consuming data in our daily lives, and the technology that has evolved around us in the recent years, where everything is connected and gadgets and IoT devices exchange bits of data every second to make our life easy, it is apparent that Data is the fuel for the future.

### **What is Big Data?**

Big data is a collection of structured, semi structured, and unstructured data that may be harvested for information and used in machine learning, predictive modelling, and other advanced analytics activities. (Pence, 2014). Although the concept of big data is still young, the origins of large data sets dates back to the 1960s and '70s when the world of data was just getting started with the emergence of RDBMS. A good definition is given by the McKinsey Global Institute: "Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse" (Manyika et al., 2011). According to IBM, Big Data is a high-volume, high-velocity, and high-variety information asset and it requires innovative, cost-effective, and advanced forms of processing techniques to have improved insights of the collected information and to use it in decision-making. Because of its massive volume in both unstructured and structured data forms the processing of Big Data using traditional RDBMS tools is a difficult task. So sometimes Big Data is also referred to as tools and technologies adopted by organizations to store and manage such a large volume of data. The Three V's which were initially used by McKinsey Global Institute to characterize Big Data; Volume, Variety, and Velocity (Chen et al., 2014) (Hu et al., 2014). These characteristics have now been extended to 10 V's i.e. Volume, Velocity, Variety, Veracity, Value, Vitality, Viscosity, Visualization, and Vulnerability. (Kepner et al., 2014; Saha & Srivastava, 2014) (Khan et. al., 2014). We could not deny the possibility of adding new characteristics in the future. Big data has not only triggered the emergence of new research approaches and practices, but have also nudged normative changes and sparked controversies regarding how research is ethically justified and conceptualized. Richterich(2018).

These small chunks of data we generate and encounter every time it has to travel through different stages, which we call the Big Data lifecycle. Which are as follows:

**Data Generation:** this is the source of data; this phase is where data is being created. Different tools and sensors and input methods can be employed to capture the data like we use sensors to capture meteorological data. We generate data on social media posts, blogs, videos etc. these are data generation sources. Mobile Phones, Self-driving cars, Smart home appliances, GPS, IoT devices with various sensors capture data every minute, which makes them a data generation point. (Hu et al., 2014; Kashyap, 2019) (Taleb et al., 2021)

**Data Acquisition:** This is the second phase and it is concerned with the process of finding information. This whole operation includes collection of data, transmission of data, and pre-processing of data. Because the data comes from different and diverse sources, it may have duplications, conflicts and redundancy, which is meaningless form of information and this may consume the storage space unnecessarily. So, data compression, and processing techniques are used for efficient data storage. (Chen et al., 2014) (Kashyap, 2019)

**Data Storage:** This is the third phase where the refined data sets are chopped into many small bundles of data and are stored on different computer nodes in a distributed manner. Network of these distributed nodes then creates a massive storage system which is robust, reliable, and fail proof. Traditional Database management systems like SQL are not used as they can not handle such a huge amount of data. Database Management systems like NoSQL and Hadoop HDFS by Apache prove to be best for storing and managing such bulky data and they provide flexible query and fast retrieval. (Kashyap, 2019) (Taleb et al., 2021)

**Data Analysis:** This is the Fourth Phase of Big Data Lifecycle, and this is a broad area, it involves Processing, Analytics and Visualization of data. This inspects, transforms and extracts true value from large sums of data. This is done using data mining tools and machine learning algorithms which helps to extract useful insights of the available information and to help make decision making easier. Data analytics research can be classified into six critical technical areas: structured data analytics, text analytics, multimedia analytics, web analytics, network analytics, and mobile analytics. Data scientists have the expertise to apply what is to be analyzed so they become the most valuable users of this phase.

#### **Literature Review:**

Shao and Zongsheng (2019) conducted a study on China's Renewable Marine Energy Based on Big Data and explained the current situation of Marine energy standardization and analyzed the specific problems and countermeasures in China. To collect the data using platforms are social public opinion analysis of science and technology, professional intelligence analysis technology, database technology and the public network database and found the share of renewable energy is low due to unreasonable allocation of resources.

Kalantari, Ali et. al. (2017) examined 6572 publications on big data research the data retrieved from the Web of Science Core Collection Database (WOS) for the period 1980 to 19 March 2015. Study analysed the various parameters such as Citation of the research output, trend of publication, contribution of countries, distribution of key words and author keyword.

Kuc-Czarnecka, Marta and Olczyk, Magdalena (2020) revealed in his study how ethics combined with big data covering publication between 1900 and July 2020. The data has been obtained from the WOS Database using HistCite and VOSviewer software; a total 892 research papers were evaluated. Mainly focus on the topic of ethics in Big Data, measuring number of articles per journal, average number of authors per journal, global and local number of citations.

Culminate that the topic is relatively poorly represented in the scientific literature with the relatively slow growth of interest.

Amir Gandomi, and Murtaza Haider, (2015) described big data concepts, methods and analytics. The paper presents a consolidated description of big data by integrating definitions from practitioners and academics and highlighting the fact that size is only one dimension of big data. Other dimensions, such as velocity and variety like audio video formats are equally important.

Singh, Bhupinder (2017) examined bibliometric analysis of literature published with affiliation of India in Arts & Humanities Citation Index (A&HCI) of Web of Science and found 3234 indexed publications between the years 1999-2017 with 5138 citations. The number of publications of Indian authors is less in A&HCI as compared to three other Asian countries but there has been a gradual increase since 2011. Measured the data country-wise, author-wise and citation-wise and concluded even in the current age of information technology and proliferation of online journals and research databases, a larger chunk of Indian journals in the field of arts and humanities are still published only in print form which remain outside the reach of computerized research databases and indexes.

Liao Huchang et.al. (2018) has done bibliometric analysis and visualization on medical big data research. The data were downloaded from the Science Citation Index Expanded and the Social Science Citation Index databases from WOS and analysed the data with GraphPad Prism 5, VOSviewer and CiteSpace softwares to know the top trends, top institution, citations H Index., distribution of keyword, co authorship status, most cited papers, The study outcomes are publications fluctuated at low level during the initial periods of 1990s and the first decade of the 21st century, medical care is moving from a disease-centered model towards a patient-centered model and cooperation among multiple authors is widespread.

### **Objectives:**

The following criteria has been used for research output of Big Data analytics

- To expound the growth of Indian Publication and contribution of other countries.
- To identify the most prolific author and investigate the pattern of authorship.
- To know the degree of collaboration.
- To examine the source types of the publications
- To identify the most prolific collaborating organization
- To explore trends in big data analytics
- To examine identify highly cited publication and distribution of citations
- To identify the top co authorship country

**Methodology:** The data has been downloaded from Scopus database on 3<sup>rd</sup> June 2021 the term selected big data Analytics in title, abstract and keyword. Further it is limited by country “India”

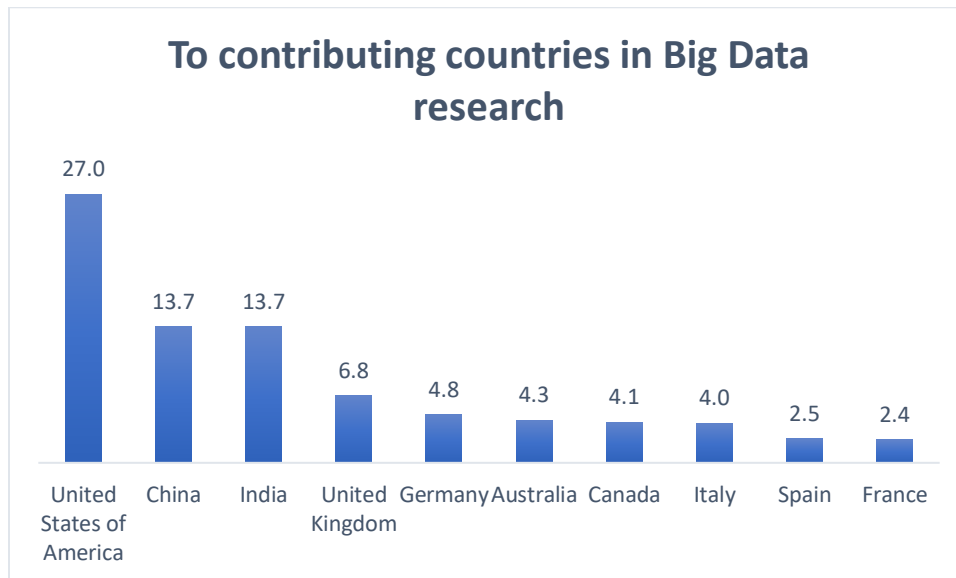
and the covering period is 2011 to 2020. We have obtained a total 2304 publication from India. VOSviewer was used for keywords, Co-authorship countries and organization analysis.

### Top ten Country share publications

Sl.no.	Country	article	%
1	United States	4554	27.03152
2	China	2312	13.72351
3	India	2304	13.67603
4	United Kingdom	1145	6.796462
5	Germany	813	4.825785
6	Australia	722	4.285629
7	Canada	692	4.107556
8	Italy	672	3.988841
9	Spain	416	2.469282
10	France	396	2.350567
	<b>N=16847</b>	<b>14026</b>	<b>83.25518</b>

### Year wise contribution

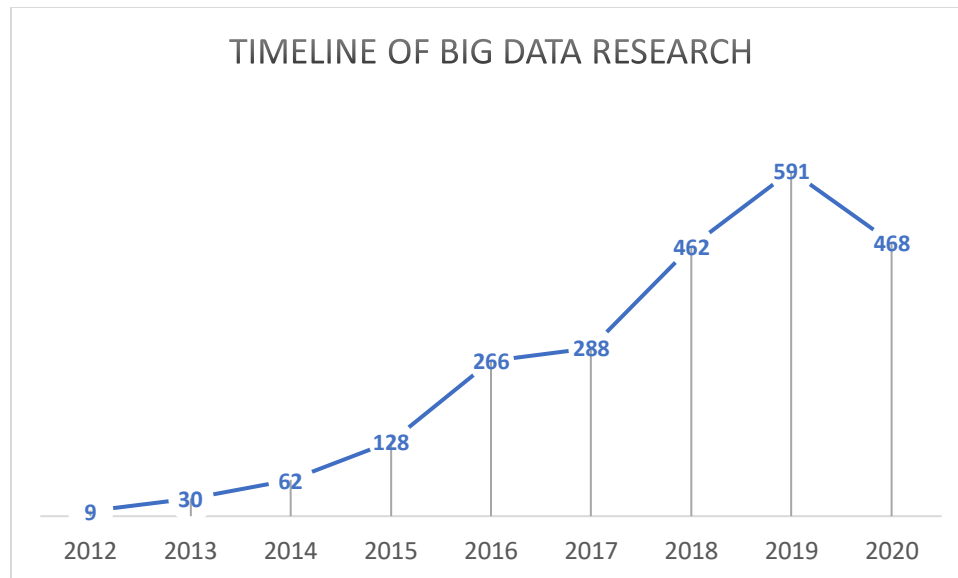
The above tables give a detailed view of the literature contributed by top ten countries in the field of Library and Information Science research. A Total of 14026 documents were published within ten years and the above table represents the top ten countries that contributed more than 83 % of the total contributions where India ranked third in the top ten countries. India Despite being under-privileged in terms of funding and infrastructure for the libraries, still contributes more than other advanced countries, like the United Kingdom, China, Canada, Australia, and Spain. However, the data pertaining to India, included in our survey is from 2012 and above as there are no articles or documents available on the specific topic before 2012.



The above graph shows that the USA is at the top position in research on Big Data Analytics worldwide, with a whopping 27% total share. India is at third rank with its research literature of 13.7% in Big Data Research, which makes us the third largest after China which is producing 13.7% almost the same research as ours. United Kingdom has been working in this research area and has been contributing 6.8% of Big Data Research and after that Germany scores 5th place with 4.8% of its research contribution. Many other developed countries like Australia and Canada are doing research in Big Data and they publish 4.3% and 4.1% of total LIS literature, respectively. Whereas Italy, Spain and France despite being developed European nations, make 4.0%, 2.5% and then 2.4% of the research submissions each respectively, and out of these top ten most contributing countries.

Sl.no.	Year	Article	%
1	2012	9	0.390625
2	2013	30	1.302083
3	2014	62	2.690972
4	2015	128	5.555556
5	2016	266	11.54514
6	2017	288	12.5
7	2018	462	20.05208
8	2019	591	25.65104
9	2020	468	20.3125
	<b>N=2304</b>	2304	100

The publication trend in Big Data Research in India shows that there has been consistent growth in India's contribution in this field. In the last ten years, a total of 2304 articles have been published. It is evident in the below given chart that the year 2017, 2018 and 2019 witnessed rapid growth in Big Data Research literature published in SCOPUS indexed journals. Though, During the Covid pandemic of 2020 the publication and contribution rate by Indian researchers has decreased.

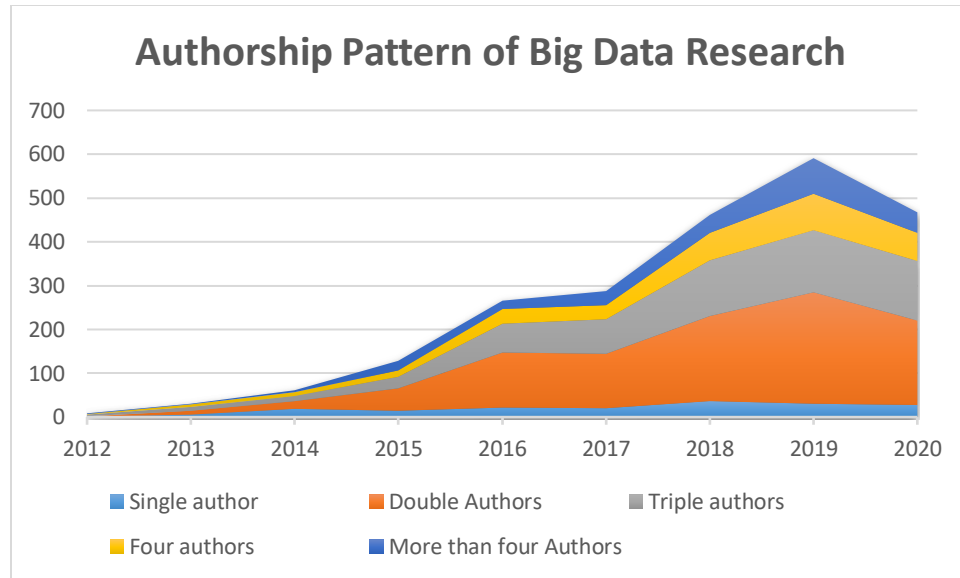


### Authorship pattern

Year	Single author	Double Authors	Triple authors	Four authors	More than four Authors	Total articles
2012	1	1	4	1	2	9
2013	6	8	10	5	1	30
2014	19	17	12	9	5	62
2015	15	50	27	14	22	128
2016	22	125	66	34	19	266
2017	21	123	80	32	32	288
2018	37	193	128	63	41	462
2019	31	254	142	83	81	591
2020	28	192	136	64	48	468
						2304

If we look at the authorship pattern in publication, we can recognize the trend that most of the publications are double authored and the year 2019 witnessed the most publications in double author category, with 254 publications in that year, whereas the number of double author papers in Big data research in the year 2018 were 193. The Trend shows consistent and almost similar patterns in single author and triple author article contributions throughout this time. The pattern shows that most of the publications are double authors makes a total of 963, double author contributions in Big Data Research, and then a total 180 single author papers were published in the last ten years. 605 Triple author contributions were made, which draws a close trend line with double author contributions.





#### Analysis of single v/s multiple authors output in Big Data Research:

The degree of collaboration is defined as the ratio of collaborative publication and total number of publications. This formula was suggested by Dr K Subramanyam in the year 1983 to calculate the authorship collaborations among Single and multi-author.

$$C = \frac{N_m}{N_m + N_s}$$

Here, C refers the degree of Author Collaboration,

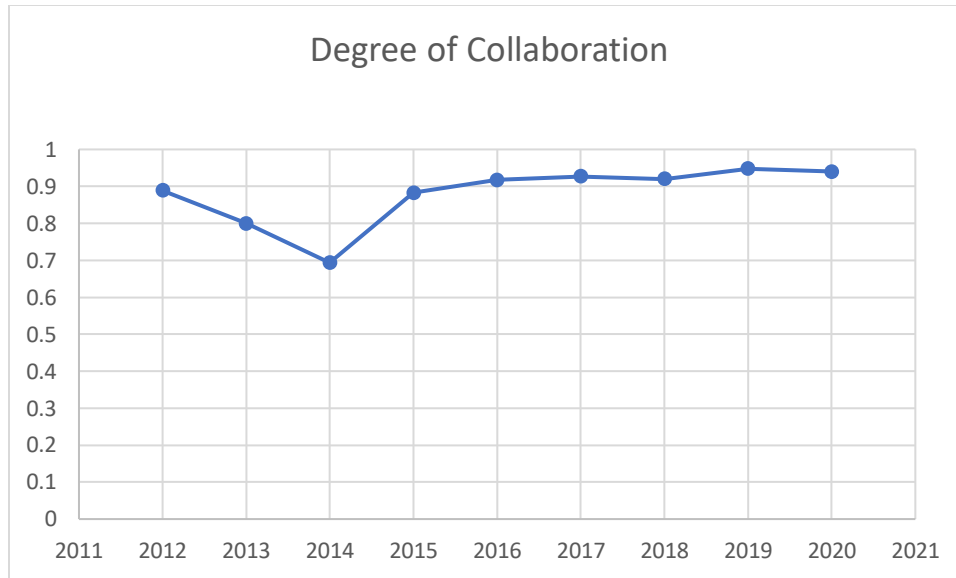
$N_m$  refers to multiple authors in a research paper, and

$N_s$  represent the single authorship.

#### Single V/S multiple authors

Year	Single authors	Multiple Authors	Degree of Collaboration
2012	1	8	0.888889
2013	6	24	0.8
2014	19	43	0.693548
2015	15	113	0.882813
2016	22	244	0.917293
2017	21	267	0.927083
2018	37	425	0.919913
2019	31	560	0.947547
2020	28	440	0.940171
Total	180	2124	0.921875

The degree of collaboration is calculated, and the value is shown in the above table. That shows the degree of Collaborative research increased during the period in the year 2015 it was 0.88, it was just after it witnessed the all-time lows of 0.69 in the year 2014. But recently in the year 2020 it was 0.92.



The evolution and succession is apparent in the above graph.

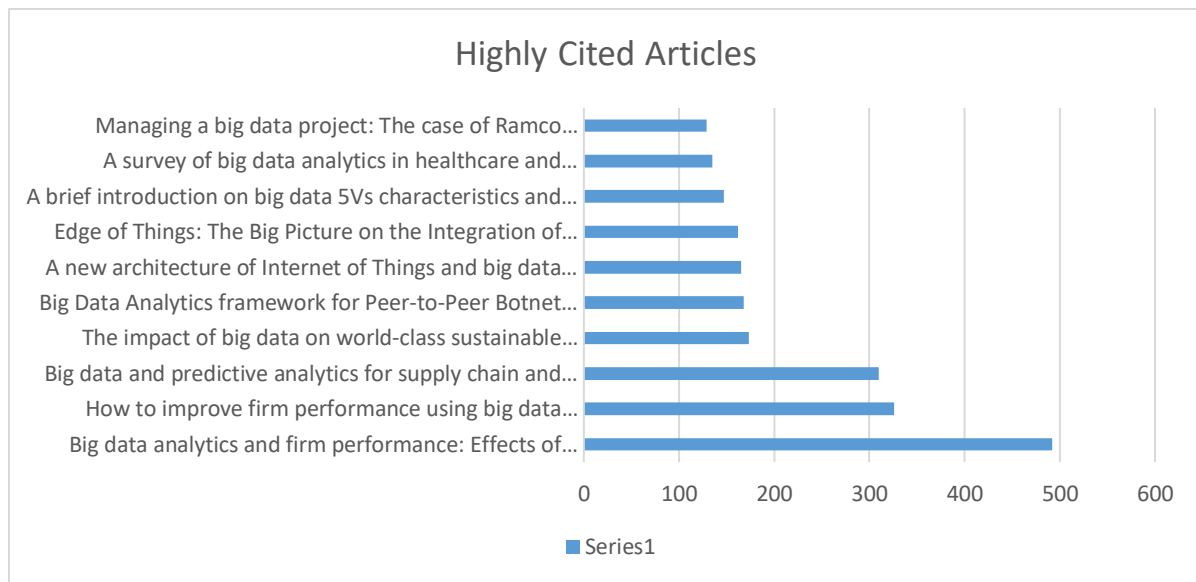
### Top ten highly cited articles

Fooladi M, et al. (2013) the citation index, as a type of bibliometric method, shows the number of times an article has been used by other papers. Wohlin C. (2005) highly cited papers have a greater chance of visibility, thus attracting greater attention among researchers.

Sl.no.	Title	citation	%
1	Big data analytics and firm performance: Effects of dynamic capabilities	492	3.605716
2	How to improve firm performance using big data analytics capability and business strategy alignment?	326	2.389154
3	Big data and predictive analytics for supply chain and organizational performance	310	2.271894
4	The impact of big data on world-class sustainable manufacturing	173	1.267864
5	Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests	168	1.23122
6	A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system	165	1.209234
7	Edge of Things: The Big Picture on the Integration of Edge, IoT and the Cloud in a Distributed Computing Environment	162	1.187248
8	A brief introduction on big data 5Vs characteristics and	147	1.077318

	hadoop technology		
9	A survey of big data analytics in healthcare and government	135	0.989373
10	Managing a big data project: The case of Ramco cements limited	129	0.945401
	N=13645	2207	16.17442

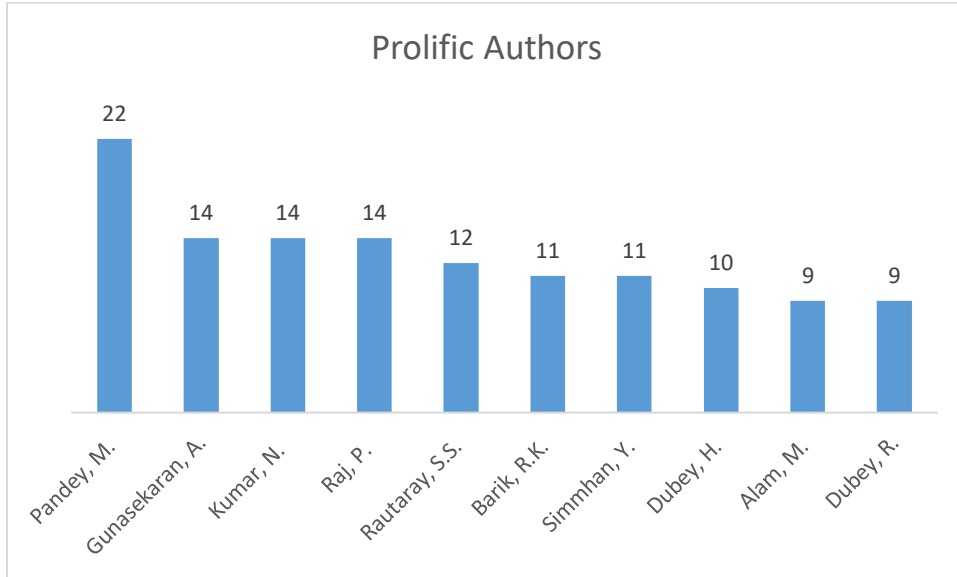
The above table shows the most cited Indian papers in Big Data research, which consist of topics like Performance improvement of systems, Sustainable manufacturing and development, and its integration and uses in Computer and Medical studies.



### Prolific authors

Sl.no.	Author Name	No of articles
1	Pandey, M.	22
2	Gunasekaran, A.	14
3	Kumar, N.	14
4	Raj, P.	14
5	Rautaray, S.S.	12
6	Barik, R.K.	11
7	Simmhan, Y.	11
8	Dubey, H.	10
9	Alam, M.	9
10	Dubey, R.	9

Here, the above table gives the details of top authors and their number of contributions in Big Data Research.

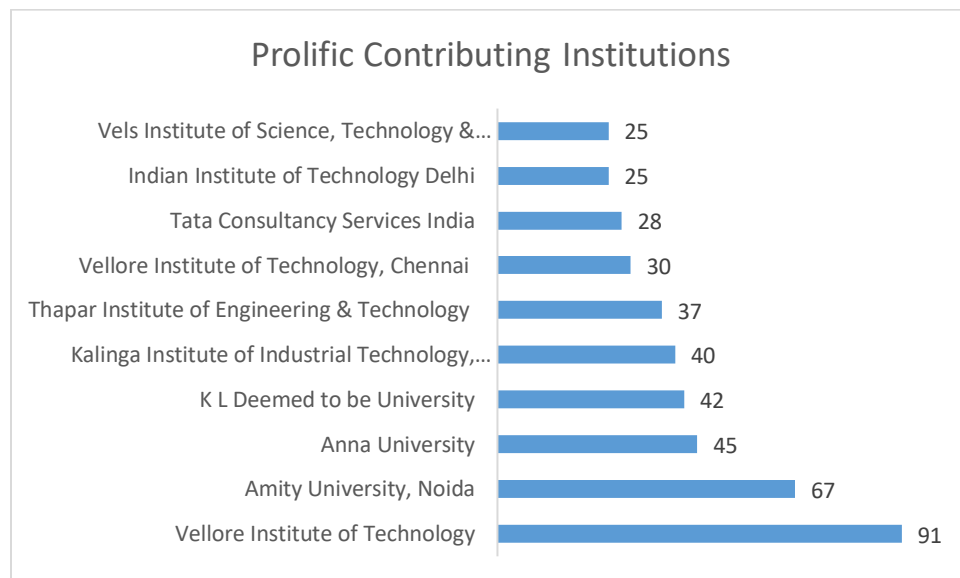


### Prolific Organization

	Organisation	No of article	%
1	Vellore Institute of Technology	91	3.949653
2	Amity University, Noida	67	2.907986
3	Anna University	45	1.953125
4	K L Deemed to be University	42	1.822917
5	Kalinga Institute of Industrial Technology, Bhubaneswar	40	1.736111
6	Thapar Institute of Engineering & Technology	37	1.605903
7	Vellore Institute of Technology, Chennai	30	1.302083
8	Tata Consultancy Services India	28	1.215278
9	Indian Institute of Technology Delhi	25	1.085069
10	Vels Institute of Science, Technology & Advanced Studies	25	1.085069
	N=2304	430	18.66319

Out of Total 2304 Scopus indexed articles the most of the number, 91 to be exact are published from Vellore Institute of Technology making their contribution to 3.9%. Amity University is one of the premier institutes in North India has made 2.9% share in Big Data Research with their 67

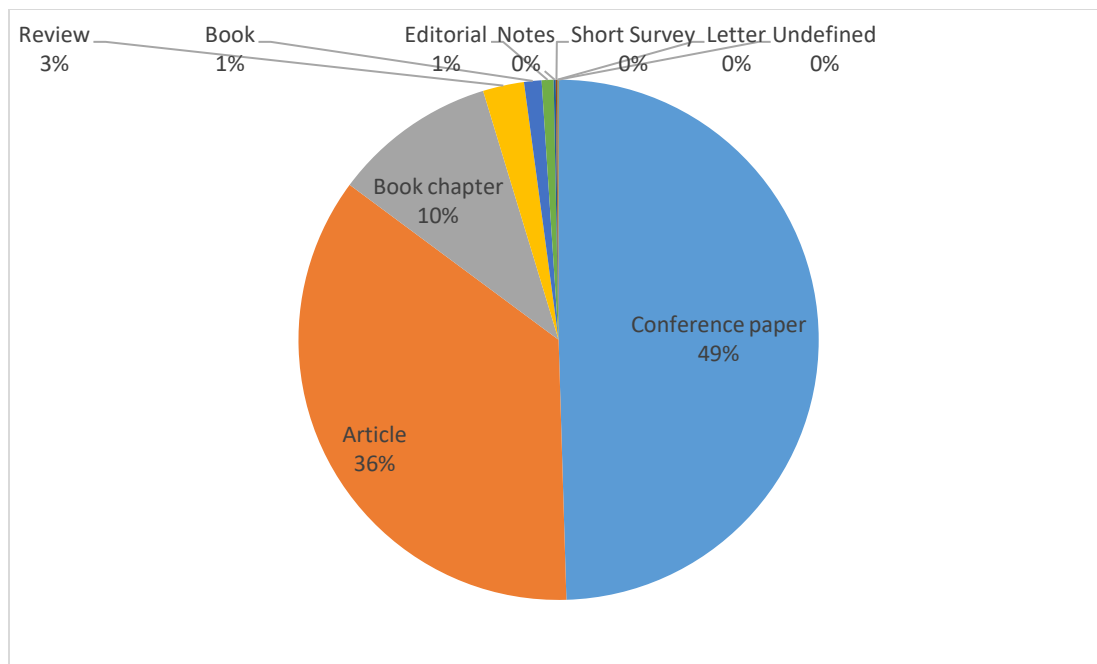
publications in the last 10 years. Anna university being at the 1.9% has made contributions of 45 research papers.



### Publication in Document Type

S No	Types of Documents	No of publications	%
1.	Conference paper	1141	49.52257
2.	Article	821	35.63368
3.	Book chapter	234	10.15625
4.	Review	59	2.560764
5.	Book	25	1.085069
6.	Editorial	17	0.737847
7.	Notes	3	0.130208
8.	Short Survey	2	0.086806
9.	Letter	1	0.043403
10.	Undefined	1	0.043403
	Total	2304	100

When we explore the distribution of the Research work published in the field of Big Data research, we can see that much of the literature is published in the form of conference papers, which counts 49.5 % of the total contribution. The researchers have published 35.6% of their work in various Journal Articles. 10% of the research work is published in the form of Book Chapters.

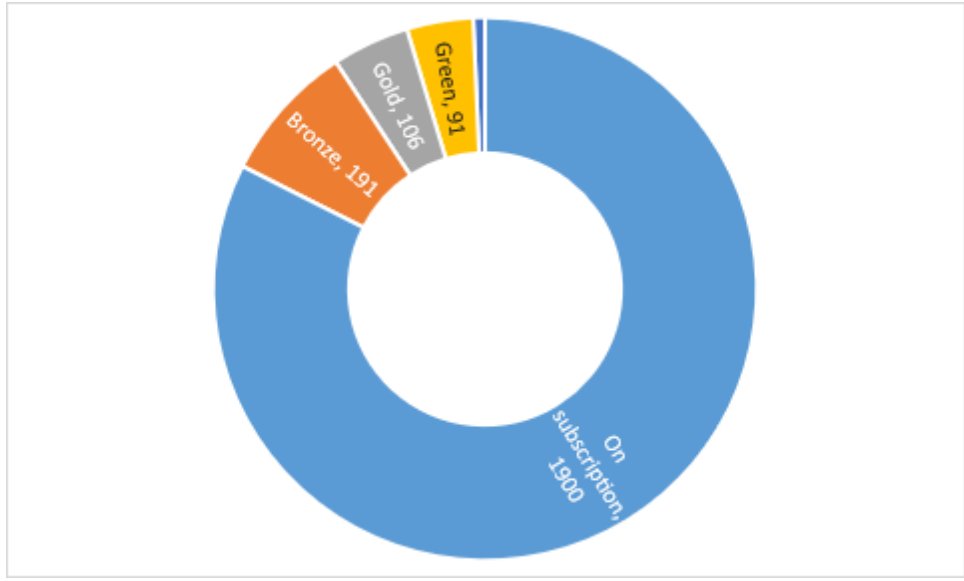


It is also clear in the above pie chart that most of the research work in Big Data Research is confined in Conference papers, Journal Articles and then Book Chapters majorly. There are very few Books, Editorials and Reviews published in the research area of Big Data. Only 25 books were published in the last ten years on the topic of Big Data research, which only comes to 1% of the total published work in the field.

### Access Type

Sl.no.	Types of Access	No. of Publications	%
1	On subscription	1900	82.46528
2	Gold	106	4.600694
3	Hybrid Gold	16	0.694444
4	Bronze	191	8.289931
5	Green	91	3.949653
		2304	100

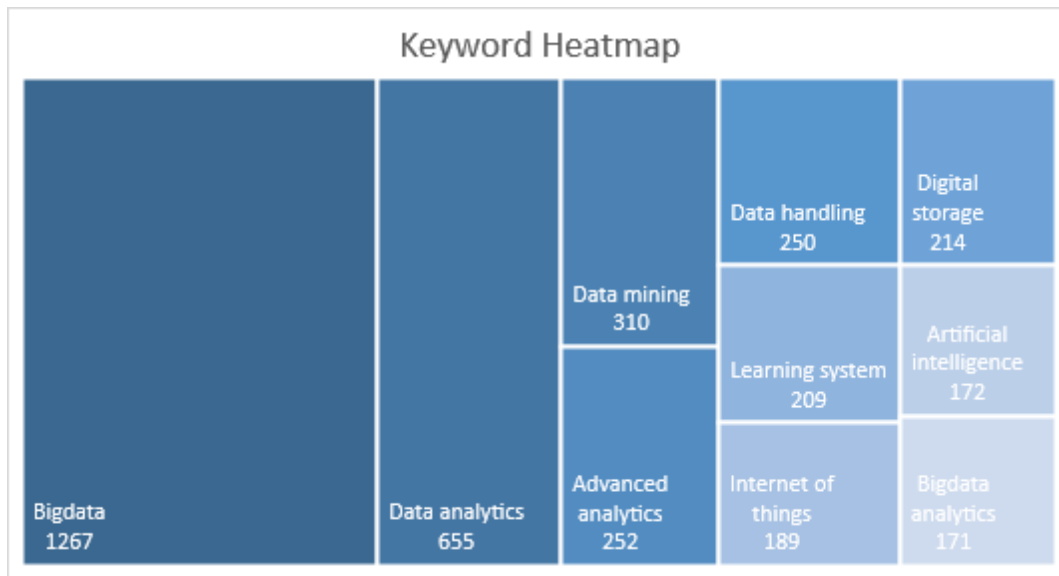
Most of the research work in Big Data was published in the journals which are on subscription-based models. And only 3% articles were published in the Green Access model.



**Related subject Trends in Big data analytics**

Sl.no.	Keywords N=7986	No of time occurred	%
1	Big data	1267	15.86526
2	Data analytics	655	8.201853
3	Data mining	310	3.881793
4	Advanced analytics	252	3.155522
5	Data handling	250	3.130478
6	Digital storage	214	2.679689
7	Learning system	209	2.61708
8	Internet of things	189	2.366642
9	Artificial intelligence	172	2.153769
10	Big data analytics	171	2.141247
		<b>3689</b>	<b>46.19334</b>

The table shows the distribution of keywords used by the Authors in their research work and the number of occurrences they have made in their work. Since the topic is Big Data Research, it is obvious for this keyword to make maximum occurrences 1267 times. But the distribution of keywords also shows the specific areas which the research papers covered like Data Analytics, Data Mining and Data handling are some of the most used keywords related to Big Data Research. This shows that the research is not only being conducted in Big Data Analytics, but also Curation, Storage and Handling of the Data.



Although Big Data in itself is a huge domain and its uses are ubiquitous in the modern world, the heatmap above shows the keywords which came into occurrence in relation to Big Data Research. The above given heat map shows how much weight each keyword holds in the total research work analyzed. The research is also being conducted in topics like IoT, Artificial Intelligence and Learning Systems, where the use of Big Data Research is presented.

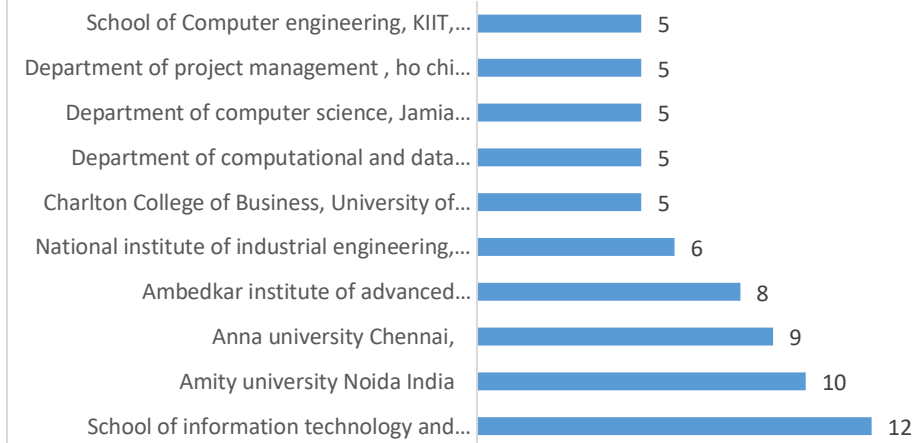
#### Top ten Co-authorship organizations:

Sl.no.	Total organizations N=3794	No of article collaborated
1	School of information technology and engineering, Vellore institute of technology, Vellore	12
2	Amity university Noida India	10
3	Anna university Chennai,	9
4	Ambedkar institute of advanced communication technologies and research , New Delhi	8
5	National institute of industrial engineering, Mumbai	6
6	Charlton College of Business, University of Massachusetts Dartmouth, North Dartmouth, US	5
7	Department of computational and data science, IISc	5
8	Department of computer science, Jamia Millia Islamia	5
9	Department of project management , ho chi minh city open university, Vietnam	5
10	School of Computer engineering, KIIT, university	5
		70

The total 3794 organizations have published 2304 articles. The above institutions have collaborated with minimum five articles. The others have not come into the list.



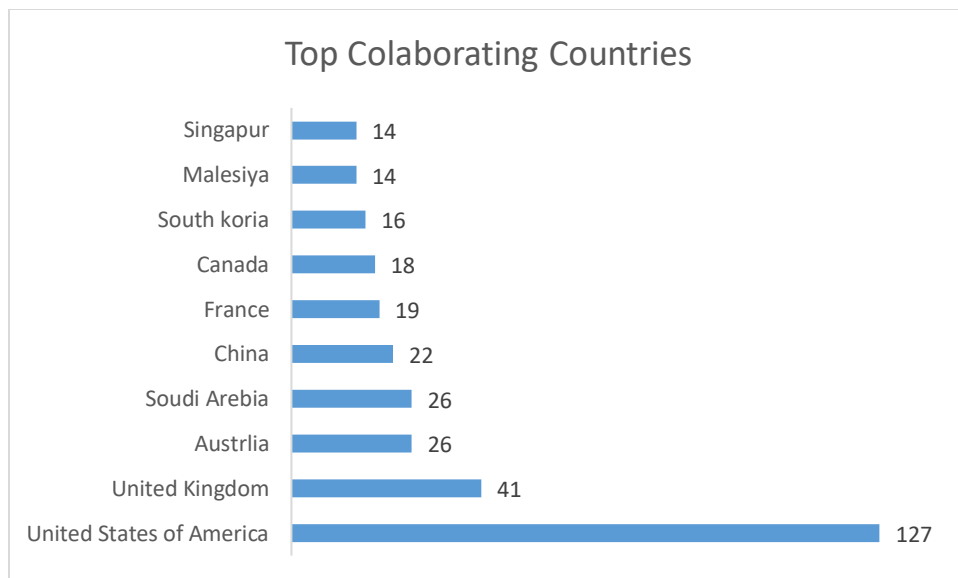
### Top Collaborating Institutions



### Top Ten Co-authorship country:

Sl.no.	Total country N=65	No of article
1	US	127
2	UK	41
3	Australia	26
4	Soudi Arabia	26
5	China	22
6	France	19
7	Canada	18
8	South korea	16
9	Malesiya	14
10	Singapur	14

The above table represents the top ten countries with which Indian research has collaborated in the publication of research articles. Indian authors have collaborated with authors of 65 countries for the publication of 323 research outputs. The distribution is more perceptible in the chart below.



The above graph accurately represents the collaboration pattern of Indian researchers with researchers from top 10 countries in the study of Big Data Research. Out of 323 collaborative publications, most of the literature was published in collaboration with the United States of America. The United Kingdom becomes the second country which has collaborated with Indian researchers with 41 joint publications. Australia and Saudi Arabia are both at the same pace becoming third co contributors with which Indian Big Data researchers have collaborated and about a total of 26 articles were published. The authors from other countries, like China, France, Canada, South Korea, Malesiya and Singapore have also made joint contributions in Big Data research published by Indian Authors.

## Conclusion

The study shows the contribution of Indian authors in the field of Big Data Research over the last 10 years and Big Data research has shown growth during the period of 2011-2020. The study shows the continuation growth of LIS research but in the year 2020 the publication rate is decreased, but it was still better than 2014 when it was its all-time low. The study shows authorship patterns whether it is single, double, or multiple. Top ten highly cited Articles, top ten journals for publication, which has been the choice of researchers to publish their work. Although every paper, every author, and every Institution has made a significant impact in knowledge building in Big Data Research and even the smallest contribution is acknowledged by the researchers' community, but we also shed some light on the Prolific authors and top contributing Institutions in Big Data Research. The paper also discussed the Collaboration patterns of Indian Researchers with other leading countries in the field.

## References

- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>
- Fooladi M, et al. Do criticisms overcome the praises of journal impact factor? *Asian Soc Sci*. 2013;9(5):176–82
- Gandomi, Amir and Haider, Murtaza (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access*, 2, 652–687. <https://doi.org/10.1109/ACCESS.2014.2332453>
- K. Subramanyan (1983) Bibliometric studies of research collaboration : a review. *Journal of Information Science*, 6 (1), 33-38
- Kalantari, Ali et.al.(2017) A bibliometric approach to tracking big data research trends. *Journal of Big Data* 4, 30 <https://doi.org/10.1186/s40537-017-0088-1>
- Kashyap, R. (2019). Big Data Analytics Challenges and Solutions. *Big Data Analytics for Intelligent Healthcare Management*, 19–41. <https://doi.org/10.1016/b978-0-12-818146-1.00002-7>
- Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., & Cunningham, R. K. (2014). Computing on masked data: a high performance method for improving big data veracity. *2014 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–6. <https://doi.org/10.1109/HPEC.2014.7040946>
- Khan, M. A., Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, 1–5. <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Kuc-Czarnecka, M., Olczyk, M.(2020). How ethics combine with big data: a bibliometric analysis. *Humanities and Social Science Communication* 7, 137 . <https://doi.org/10.1057/s41599-020-00638-0>
- Liao Huchang et.al.(2018) A Bibliometric Analysis and Visualization of Medical Big Data Research. *Sustainability*, 10 (1), 166; <https://doi.org/10.3390/su10010166>
- Manyika, J., Chui, M., & Brown, B. (2011, May 1). *Big data: The next frontier for innovation, competition, and productivity* | McKinsey. McKinsey Digital. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>

Mavuduru, A. (2020). *Is Data Really the New Oil in the 21st Century? | Towards Data Science*. Towards Data Science. <https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-century-17d014811b88>

Palmer, M. (2006). *ANA Marketing Maestros: Data is the New Oil*. [https://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](https://ana.blogs.com/maestros/2006/11/data_is_the_new.html)

Pence, H. E. (2014). What is Big Data and Why is it Important? *Journal of Educational Technology Systems*, 43(2), 159–171. <https://doi.org/10.2190/et.43.2.d>

Richterich, A. (2018). The Big Data Agenda: Data Ethics and Critical Data Studies. Pp. 1–14. London: University of Westminster Press. DOI: <https://doi.org/10.16997/book14.a>. License: CC-BY-NC-ND 4.0

Saha, B., & Srivastava, D. (2014). Data quality: The other face of Big Data. *2014 IEEE 30th International Conference on Data Engineering*, 1294–1297. <https://doi.org/10.1109/ICDE.2014.6816764>

Shao, M., & Wu, Z. (2019). China's Renewable Marine Energy Based on Big Data. *Journal of Coastal Research*, 597–600. <https://www.jstor.org/stable/26853324>

Singh, Bhupinder, Bawa, S K and Singh, Amandeep (2017). Contribution of Indian Authors in Web of Science: Bibliometric Analysis of Arts & Humanities Citation Index (A&HCI). *International Journal of Library & Information Science*, 6(5) 10–16. available at <http://iaeme.com/Home/issue/IJLIS?Volume=6&Issue=5>

Taleb, I., Serhani, M. A., Bouhaddioui, C., & Dssouli, R. (2021). Big data quality framework: a holistic approach to continuous quality management. *Journal of Big Data* 8,(1). Springer International Publishing. <https://doi.org/10.1186/s40537-021-00468-0>

Wohlin C. (2005) An analysis of the most cited articles in software engineering journals:1999. *Information and Software Technology*,47(15), 957–64 <http://urn.kb.se/resolve?urn=urn:nbn:se:bth-8796>

Xu, T.; Yong, G., and Sarkis, J., (2018). Trends and features of embodied flows associated with international trade based on bibliometric analysis. *Resources Conservation & Recycling*, 131, 148-157. ISSN 0921-3449 <https://doi.org/10.1016/j.resconrec.2018.01.002>