

# A BIBLIOMETRIC ANALYSIS OF THE GLOBAL RESEARCH ON BIG DATA FOR SUSTAINABLE DEVELOPMENT

**Andreea Corina DANILA**

“Ștefan cel Mare” University of Suceava, Romania  
[andreea.danila@usm.ro](mailto:andreea.danila@usm.ro)

**Carmen Eugenia NASTASE**

“Ștefan cel Mare” University of Suceava, Romania  
[carmen.nastase@usm.ro](mailto:carmen.nastase@usm.ro)

Received 28 March 2021; Accepted 29 May 2021

## **Abstract:**

*There is a growing interest in the literature on the theme of big data and sustainable development. As the volume of data is increasing exponentially so does the need to analyse, understand and use this data in the context of UN 2030 Agenda for Sustainable Development Goals. The data revolution, the open data movement, the rise of Artificial Intelligence and the Internet of Things have the power to transform society through integrated actions in order to tackle the environmental, economic and social challenges that we are facing today. Despite all these opportunities, there is still a lack of knowledge on how to make the best use of all existing big data and how it can be mapped in research areas so the public can benefit. Therefore, this paper proposes a bibliometric analysis on the subject of big data for sustainable development with the aim to assist in the development of future research papers. To this end, journals indexed in Scopus database have been used. We extracted a total of 281 papers that were further analysed using the following criteria: publication year, affiliation, country, funding sponsor, type of research document, subject area and most identified keywords. The results of the review serves as the basis for determining possible future research projects, raising awareness at the country level as to the lack of funding towards this domain and the achievement of SDG sustained by the UN.*

**Key words:** big data sustainable development bibliometric analysis

**JEL classification:** Q01, Q54, Q55, Q56

## **1. INTRODUCTION**

We are living in the dawn of the fourth scientific revolution marked not only by the development of big data science and analytics but also by the mainstream adoption of advanced technologies. As the world becomes more and more digitalised, we generate on a daily basis a staggering amount of heterogeneous, structured and unstructured data that proves to be difficult to process using traditional database and software techniques. The IDC (International Data Corporation) estimates that by 2025 the global datasphere would reach 175 zettabytes ( $10^{21}$  bytes or a billion terabytes). It is no surprise that the concept of Big Data has gathered much attention in the public discourse in the last years ranging from private sector, government, international institutions and academia. While there are many definitions of the term, there is a general consensus that Big Data can be defined by the 3 Vs: volume (measured in gigabytes, terabytes, etc), velocity (the speed at which data is generated and accumulated) and variety (the massive array of data types like text, video, sound, photos). This means that systems from all industries and humans are generating data as part of their daily activities.

The substantial contribution and potential of Big Data to the international development has generated the concept of “Big Data for Sustainable Development” first coined by UN Global Pulse, an initiative launched by UN Secretary – General in 2009. The ambition was to leverage the fast development in generation, collection and analysis of digital data in order to help decision makers to understand, prevent and tackles social, economic or environmental crises with huge impact on vulnerable population. By this token, Big Data has received significant attention in the recent years to whether it can contribute to the achievement of the SDG (Sustainable Development Goals).

There are many ways that countries can make use of data to measure, monitor and improve their progress towards accomplishing the SDG and ensure accountability. As this is a new concept that merge 2 different domains, data science and sustainable development it is interesting to analyse the state of the research and the possible gaps that need to be filled in. To this extent a bibliometric analysis represents a useful method to objectively measure current state of research on a certain subject and its international scientific influence as an aspect of scientific quality . However, to our knowledge, there is no previous evidence of a global bibliometric analysis on Big Data and Sustainable Development. Our study aims to assess the quantity and quality of worldwide research in the selected field.

## **2. METHODOLOGY**

The methodology used for the systematic review of scientific articles on big data in the context of sustainable development included 2 parts. The first part included a selection of articles published in indexed journals in the Scopus database. The analysis in the second part was constructed based on the results obtained in the first step and included metrics related to a) number of articles published by years, b) institution affiliation, c) country based authors, d) funding sponsor, e) type of article, f) domain area and g) word cloud that shows the most used words in the article.

### **2.1. IDENTIFICATION AND SELECTION OF SCIENTIFIC PAPERS**

The first part of the analysis was performed based on the PRISMA protocol. Scopus was used as the database for screening scientific papers based on three criteria: a) the keywords “ big data sustainable development”, b) publication stage was limited to final and c) English as the language of the paper. No time range has been selected. As this is a fairly new research topic we didn't consider it needed a limitation in time. The first step resulted in 742 documents.

Step 2 consisted in sorting and selecting the articles. For this purpose 2 filters were used: a) only articles and conference papers have been kept in the selection: b) articles published in journals that addressed one of the subject areas related to sustainable development (Energy, Environmental Sciences, Social Sciences, Agriculture and Biological Sciences and Economics Econometrics and Finance).

After all filters were applied, 281 articles have been generated as eligible to be considered in the present research paper.

### **2.2. ANALYSIS OF PAPERS – CLASSIFICATION CRITERIA**

Step 3 of this analysis included the criteria on which these papers were classified. In total we could make our analysis based on 7 criteria. We looked at the a) year in which they were published, b) affiliation, c) country or territory from where the authors belong, d) funding sponsor; e) type of article, f) subject area and g) cloud words.

Publication year refers to the year in which the paper was finally published. This metric provides us with an insight on the evolution of the subject in the academia. It is worth mentioning that we haven't restricted the search to specific time range, focusing on the entire sea of papers published on the subject of big data and sustainable development. Affiliation criteria makes reference to the authors affiliation to research institutions that support the publication of research papers on certain domain of interests.

Country or territory refers to the geographical location where the authors or the research institutions to which they are affiliated, are based. This gives us a full picture of the geographical concentration of this particular topic of big data in the context of sustainable development.

When we are referring to the funding sponsor, we are interested to see from where the grants are coming and what institutions are actively supporting continuous research in the filed of sustainable development and big data.

Scopus database includes different types of papers ranging from articles published in journals with a focus on a certain subject like energy or environmental sciences, conferences papers presented at conferences held with a specific field of interest, article reviews and book reviews.

Subject area criteria refers to the main topic on interest for a journal in which the papers are being published. It is interesting to see the subject area most common when talking about big data and sustainable development.

**Table 1. Paper classification criteria**

<b>Criterion</b>	<b>Description</b>
Publication year	Year in which the paper was published
Affiliation	Research institutions to which the authors of the published paper are affiliated
Country/Territory	Country or territory where the institutions are based
Funding sponsor	Institutions that offered grants to the authors to conclude research on a certain topic
Type	The papers can be published as articles in Journals or as conference papers at specialised conferences.
Subject area	Subject area of the journal/conference in which the paper is published
Cloud words	Most used words in the analysed papers based on keywords indicated by the authors and the ones selected by the Scopus Database.

Scopus database includes different types of papers ranging from articles published in journals with a focus on a certain subject like energy or environmental sciences, conferences papers presented at conferences held with a specific field of interest, article reviews and book reviews.

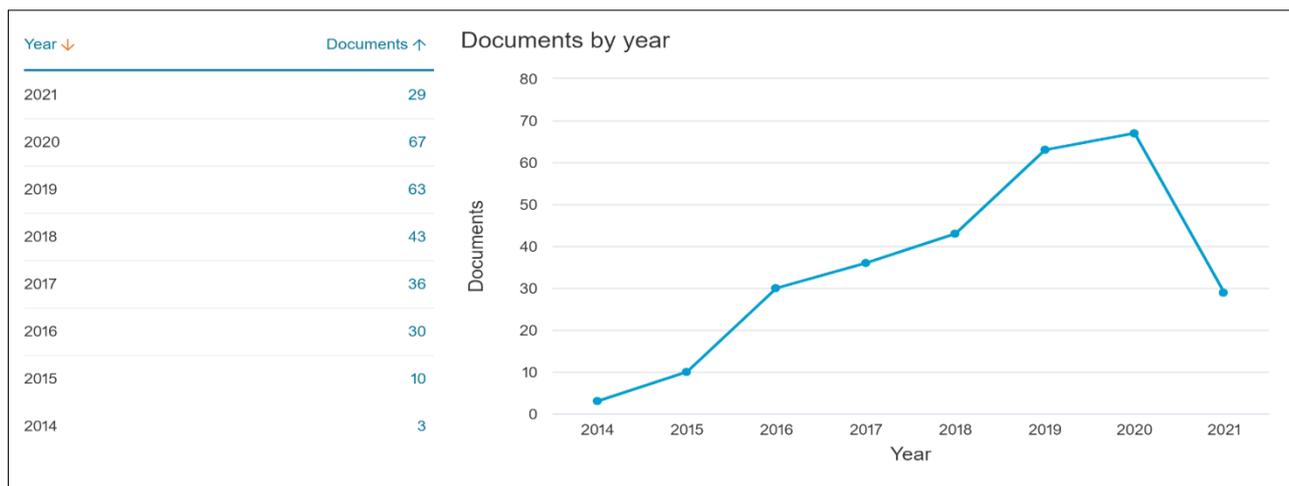
Subject area criteria refers to the main topic on interest for a journal in which the papers are being published. It is interesting to see the subject area most common when talking about big data and sustainable development.

To see what words are more prominent in the analysed papers, we have decided to make a wordle, also known as word clouds. For this metric we took as reference the authors keywords and the index keywords generated by the Scopus database. According to Scopus support centre, author keywords are words chosen by the author(s) which, in their opinion, best reflect the contents of their published document. Indexed keywords are words chosen by content suppliers and are standardized based on publicly available vocabularies. Unlike Author keywords, the Indexed keywords take into account synonyms, various spellings, and plurals. With respect to these metric used in our analysis Scopus has no influence over either Author or Indexed keywords because these are both determined by third parties. From the words database of both indicators we have removed the keywords on which the search was made “big data sustainable development” as these will appear by default and will have the highest share in the keywords section. We also removed the words “sustainability”, replaced “internet of things” with “IoT” and removed common words like “of”, “and”, “on”. Pro Word Cloud was used as a tool to create the word cloud.

### 2.3. RESULTS OF THE ANALYSIS AND DISCUSSION

Figure 1 shows the number of publications by year. As already mentioned we did not restrict our search to a specific time range. We performed our bibliometric analysis on documents published until April 2021. From the results obtained, we see that the subject is gaining increased interest in the academia from only 3 documents published in 2014 to 67 papers published or presented in conferences by the end of 2020.

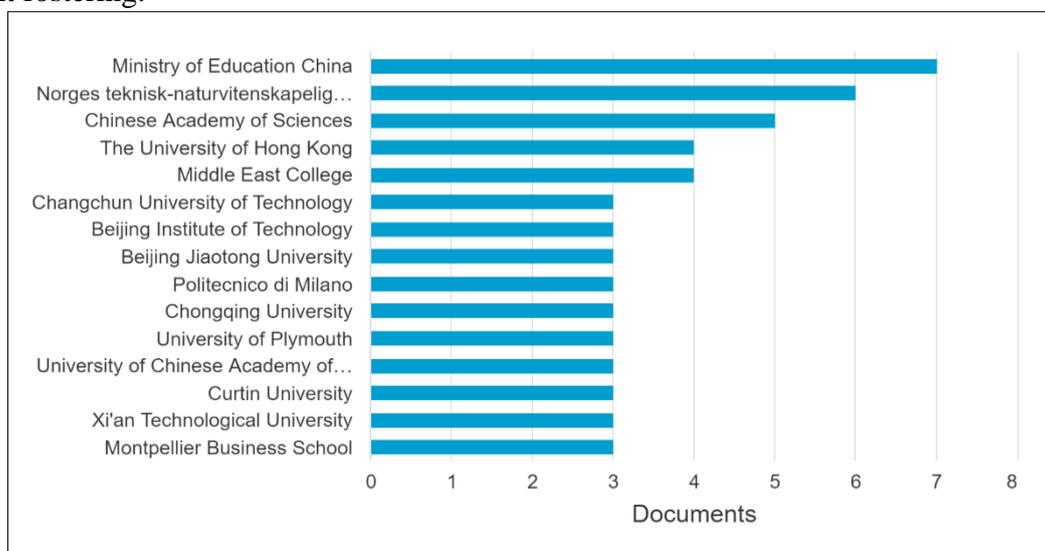
When referring to affiliation metrics, Figure 2 depicts the Institutions affiliated with the authors publishing the analyzed papers. In the top of the rank is the Chinese Ministry of Education followed by the Norwegian University of Science and Technology, Chinese Academy of Science and the University of Hong Kong.



**Figure 1. Documents published by year**

Source : Scopus database

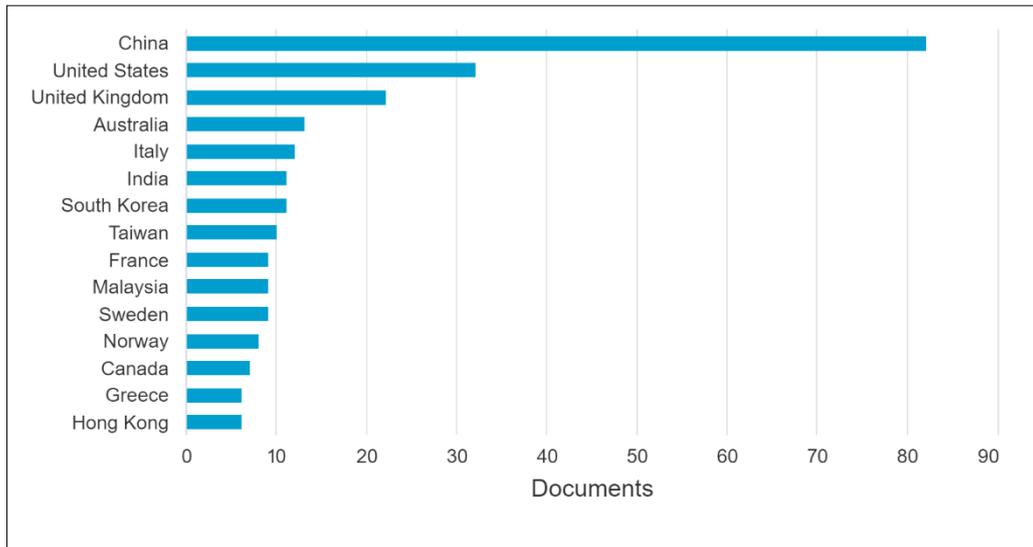
The ranking includes mostly Chinese based institutions that show an increased interest for the topic of big data in the context of sustainable development. This should come with no surprise as the Chinese province Guizhou will become in the near future the Big Data Valley for Sustainable Development. Chinese President XI Jinping announced during his UN General Assembly held on September 22, 2020 that it will sustain the development of an International Research Center to facilitate the implementation of the UN 2030 Agenda. The center will aim to build a big data technology platform through integration of research, technological innovation, capacity building and talent fostering.



**Figure 2. Documents published by affiliation**

Source : Scopus database

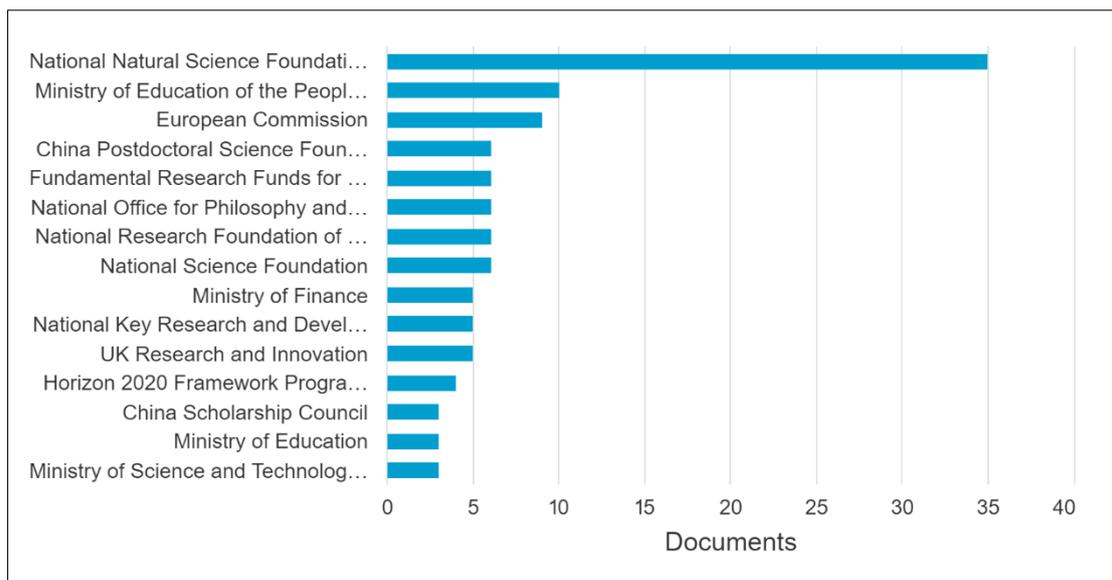
In terms of geographical location, as expected, China ranks first with 82 papers published followed by the United States with 33 articles representing a third only from the first ranked country. United Kingdom follows with 24 papers, Australia with 12 and Italy with only 14. It seems that European researchers do not manifest high interest in new ways to promote sustainable development through the use of big data. One reason would be that Europe is lacking behind US and China when it comes to innovation on data technologies but also due to its GDPR policy which impacts the collection of Big Data in general.



**Figure 3. Documents published by country or territory**

Source : Scopus database

In terms of funding the research papers, China holds the first two positions with the National Natural Science Foundation of China and the Ministry of Education of China, both accounting 45 papers in our sample papers. The European Commission ranks third with only 9 papers that treat the issue of sustainable development and new technology as big data and data science.

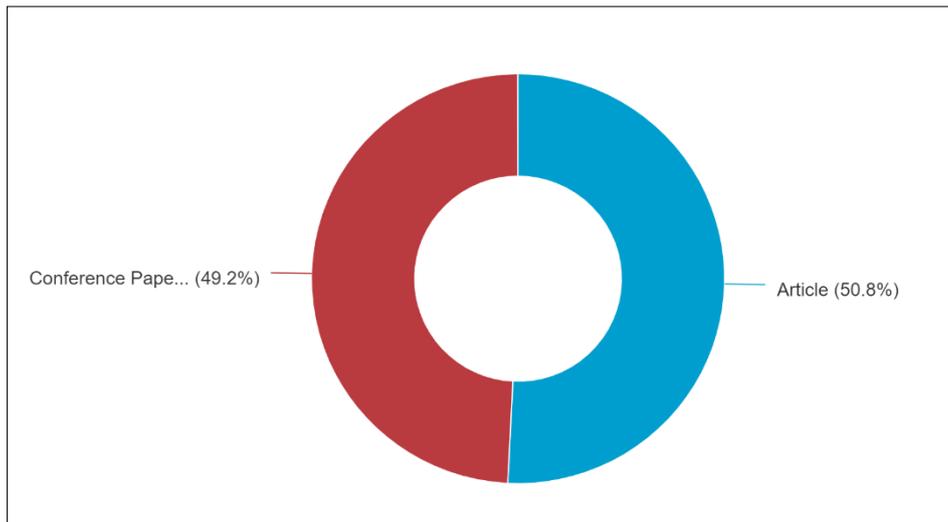


**Figure 4. Documents published by funding sponsor**

Source : Scopus database

At the beginning of the study we mentioned that we restricted our sample to only articles published in journal and conference papers. Figure 5 depicts the spread of these papers in the two

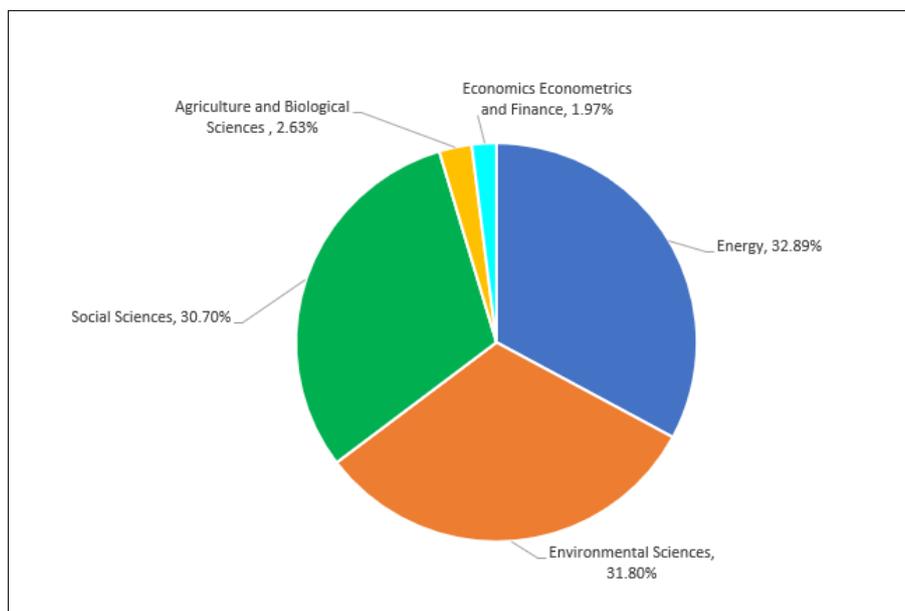
categories. Articles have a slightly higher share (50.8%) compared to conference papers. At a closer look into the data we see that most papers funded by the Chinese institutions were published as conferences papers and not in specialized journals. Paper conferences represent a good method for researchers to present new concepts and techniques in a particular field of study that is not fully developed. This is a valid assumption for the topic of big data and sustainable development considering that the publications in the field have started to increase since 2014 only. Conference papers also have the advantage that researchers can engage with leading edge scientific knowledge prior to publishing in official journals in an effort to influence the direction of the research at early stages. More than this, since conferences treat a certain domain or topic, research institutions and governments find them more appealing to discover the state of the art research and to prepare for technological changes required in the future.



**Figure 5. Documents published by type**

Source : Scopus database

Figure 6 depicts the share of documents published by subject area. The majority of the papers are published in journals or at conferences in the sphere of energy, environmental sciences and social sciences. It seems that economics and sustainable agriculture is not gaining interest in studying the potential of big data and how it can foster sustainable development.



**Figure 6. Documents published by subject area**

Source: Own elaboration using data from Scopus database



2. United Nations . Agenda 21: The United Nations Programme of Action from Rio. United Nations; New York, NY, USA: 1993.
3. World Health Organization . Health and Environment in Sustainable Development: Five Years after the Earth Summit. World Health Organization; Geneva, Switzerland: 1997.
4. United Nations Global Pulse (2013) Big Data for Development: A primer. June 2013
5. UN Global Pulse (May 2012) Big Data for Development: Challenges and Opportunities, May 2012 Hotz, Robert Lee. "The Really Smart Phone." The Wall Street Journal. 22 Apr. 2011.
6. UNDP Global Pulse. A Guide to Data Innovation for Development. From Idea to proof-of-concept. November 2017
7. UN Global Pulse & Leiden University (2016), Big Data for Development and Humanitarian Action: Towards Responsible Governance
8. UNDP Global Pulse. A Guide to Data Innovation for Development. From Idea to proof-of-concept. United Nations November 2017
9. Irizarry, R. A. (2020). The Role of Academia in Data Science Education . Harvard Data Science Review, 2(1). <https://doi.org/10.1162/99608f92.dd363929>
10. <https://www.politico.eu/article/digital-strategy-data-artificial-intelligence-europe-china-us/>
11. [https://english.cas.cn/newsroom/news/202009/t20200926\\_244297.shtml](https://english.cas.cn/newsroom/news/202009/t20200926_244297.shtml)