

BIG DATA STUDIES: A SYSTEMATIC REVIEW

Mehmet Ali Köseoglu, *The Hong Kong*, MehmetAli.Koseoglu@polyu.edu.hk
Mehmet Yildiz, *Yildirim Beyazıt University*, yildiz@ybusm.info
John P. Girard, *Middle Georgia State University*, john.girard@mga.edu

ABSTRACT

This systematic review of the progress of big data studies over the period 2009 to 2016 yielded a series of interesting findings. Of the 87 articles analyzed, all from the top information systems or top knowledge management journals, nearly half were published in 2016; more than half were published in just four journals; more than half were written by two or three authors; most were interdisciplinary in nature; and most originated from two countries.

Keywords: Big data, bibliometric analysis, journals, knowledge management, information systems

INTRODUCTION

The purpose of this article is to chronicle the progress of big data studies with a view to sparking additional research. This project was to be the first phase of a multiphase exploratory project designed to codify the state of the big data research. After the completion of the first phase, the researchers identified a shift in big data research. Nevertheless, it was felt the initial findings would be of use to others interested in the field. As a result, this article focuses on the findings that the researchers hope will provide a foundation for additional research.

The genesis of this project was the discovery of a large increase in the interest over time for “big data.” Figure 1 shows data from Google Trends based on the relative interest for the search term “big data” over the period January 2004 to June 2016 (Google, 2016). A score of 100 represents the peak in popularity of the search term, which for “big data” was February 2016.

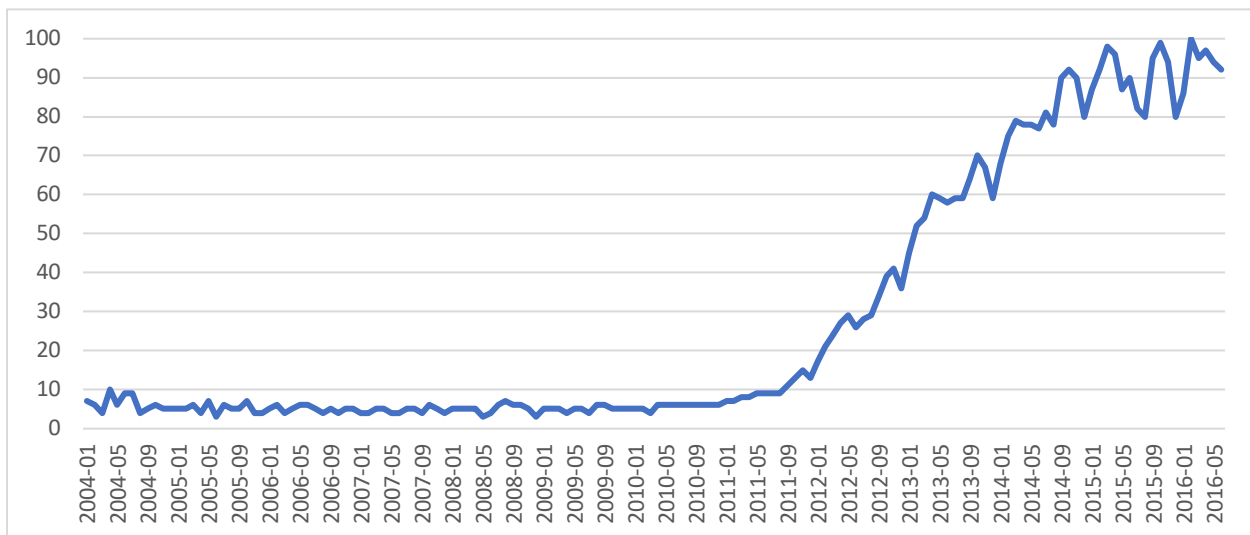


Figure 1. Worldwide Interest Over Time for Big Data (Google, 2016)

A review of the data indicates the rapid increase, not surprisingly, corresponded with a series of highly publicized cover stories about big data in a wide range of periodicals, such as *Popular Science* (Enriquez, 2011) *Harvard Business Review* (McAfee & Brynjolfsson, 2012), *Foreign Affairs* (Cukier & Mayer-Schoenberger, 2013), and *The Futurist*

(Tucker, 2013). As intriguing as this meteoric rise in the searches might be, our interest was not the popular press, but rather the academic body of knowledge, specifically peer-review journal articles.

RESEARCH METHODOLOGY

Process

The researchers followed an established methodology, modelled after other bibliometric projects (Koseoglu, Akdeve, Gedik, & Bertsch, 2015) (Ocak, Koseoglu, & Yildiz, 2017). The stages included: determine scope; finalize sampling frame; collect journal articles; complete analysis (content analysis and descriptive statistics); and interpret results.

Data Sources and Searches

Two ways were followed to conduct the searches in September 2016. First, journals publishing studies related to big data and published in English were identified. For this reason, top ranked Knowledge Management journals (Serenko & Bontis, 2013), and Information Systems journals (Lowry, et al., 2013), are considered.

The Knowledge Management Journals (listed in alphabetical order) were:

1. Electronic Journal of Knowledge Management (EJKM)
2. Interdisciplinary Journal of Information, Knowledge and Management (IJKM)
3. International Journal of Knowledge and Learning (IJKL)
4. International Journal of Knowledge Management (IJKM)
5. International Journal of Knowledge Management Studies (IJKMS)
6. International Journal of Knowledge Society Research (IJKSR)
7. International Journal of Knowledge-Based Development
8. International Journal of Knowledge-Based Organizations
9. International Journal of Knowledge, Culture and Change Management
10. International Journal of Learning and Intellectual Capital
11. Journal of Information and Knowledge Management
12. Journal of Intellectual Capital
13. Journal of Knowledge Management (JKM)
14. Journal of Knowledge Management Practice
15. Knowledge and Process Management: The Journal of Corporate Transformation
16. Knowledge Management & E-Learning: An International Journal (KMELAIJ)
17. Knowledge Management for Development Journal
18. Knowledge Management Research & Practice
19. The Learning Organization (TLO)
20. VINE: The Journal of Information and Knowledge Management Systems (VINE)

The Information Systems journals (alphabetical order) were:

1. Decision Support Systems (DSS)
2. Electronic Commerce Research and Applications (ECRA)
3. European Journal of Information Systems (EJIS)
4. Information & Management (IM)
5. International Journal of Electronic Commerce (IJECE)
6. Information Systems Frontiers (ISF)
7. Information Systems Journal (ISJ)

8. Information Systems Management (ISM)
9. Information Systems Research (ISR)
10. Information Technology and Management (ITM)
11. Journal of the Association for Information Systems (JAIS)
12. Journal of Computer Information Systems (JCIS)
13. Journal of Database Management (JDM)
14. Journal of Global Information Management
15. Journal of Information Technology (JIT)
16. Journal of Management Information Systems (JMIS)
17. Journal of Organizational Computing and Electronic Commerce (JOCEC)
18. Journal of Strategic Information Systems (JSIS)
19. Management Information System Quarterly (MIS)
20. Management Information System Quarterly Executive (MIS-E)

Second, the articles were scrutinized via keywords including big data, data scientist, MapReduce, Hadoop, and Predictive Analytics in the title, abstract, and keywords of articles in the selected journals. This technique is well established in the literature and has been used in other big data related projects (Lowry, et al., 2013).

After the researchers had collected of the relevant articles, Serenko & Bontis published a follow-up to their 2013 article (Serenko & Bontis, 2017). Their new list of Knowledge Management Journals included 19 of the 20 included in this project. In fact, 18 of the 20 journals used for this project remain in the top 20 of Serenko & Bontis' new list. One journal, the Journal of Knowledge Management Practice, appears to have ceased publication in 2013, and therefore as not on the new list. Additionally, the International Journal of Knowledge Society Research dropped from 20 on the 2013 list to 22 on the 2017 list. The researchers are confident the list of Knowledge Management Journals used for the project remains appropriate.

Inclusion Criteria

Two inclusion criteria were used in selecting studies. First, we did not place any time restrictions, and our search covered all time periods up to the end of October 2016. Second, we considered only original articles and research notes published in the given journals.

Quality and Coverage Assessment

Three authors of this paper assessed the articles individually to determine if they focused on big data research by asking, "Does the article directly or indirectly relate to big data research?" If the authors responded in the affirmative, then the article was accepted for the subsequent steps of the study. The authors ensured data validity and reliability by reaching a consensus on the articles that were selected. At the end of the step 87 studies were identified.

Variables Examined

Selected studies were reviewed to extract the following information: (a) journal title, (b) publication year, (c) themes of articles, and (d) nature of articles including article type, methods of article, research methods used, data collection method, and samples of articles as a country.

Analysis

A spreadsheet was created to examine the articles by employing information regarding variables of interest. This information was tabulated or graphitized to elucidate the state of the art of big data research.

RESULTS

Frequency of Big Data Studies

Figure 2 summarizes the number and growth of big data articles.

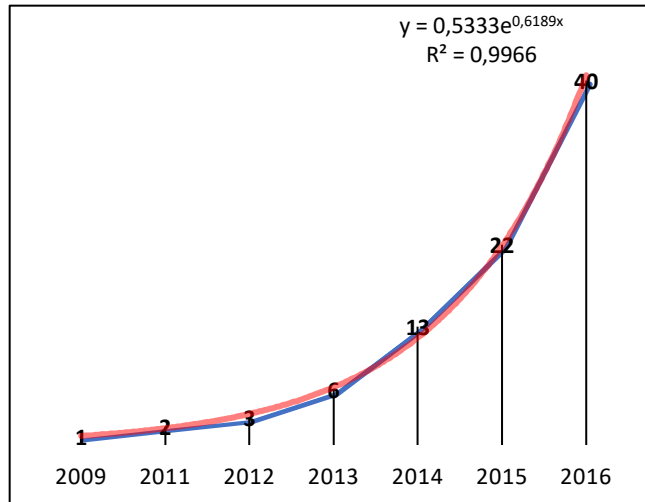


Figure 2. The Number and Growth of Big Data Articles by Year

Figure 3 demonstrates the number of big data articles by year and by journals.

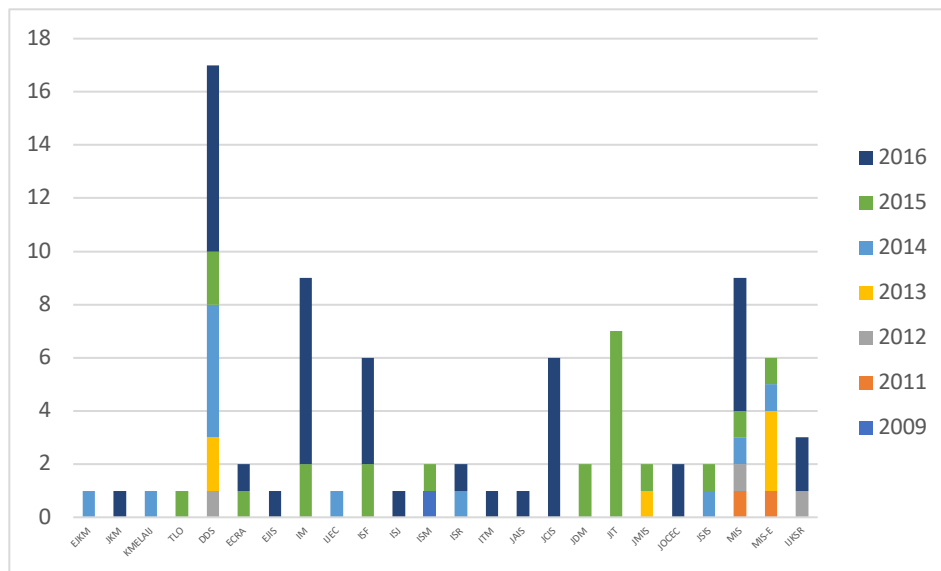


Figure 3. The Number of Big Data Articles by Year and Journals

Table 1 presents the frequency of authorship in big data studies.

Table 1. Authorship in Big Data Studies

Authorship	Frequency	Percent
Single Author	12	13.8
2 Authors	25	28.7
3 Authors	24	27.6
4 Authors	18	20.7
5 Authors	3	3.4
6 Authors	1	1.1
7 Authors	3	3.4
8 Authors	1	1.1
Total	87	100.0

Table 2 shows the collaboration types based on disciplines.

Table 2. Disciplinary Collaboration

Disciplinary Collaboration	Frequency	Percent
Interdisciplinary	44	50.6
Disciplinary	27	31.0
No collaboration	11	12.6
Unable to determine	4	4.6
Transdisciplinary	1	1.1
Total	87	100.0

Figure 4 shows the collaboration types based on institutions and countries.

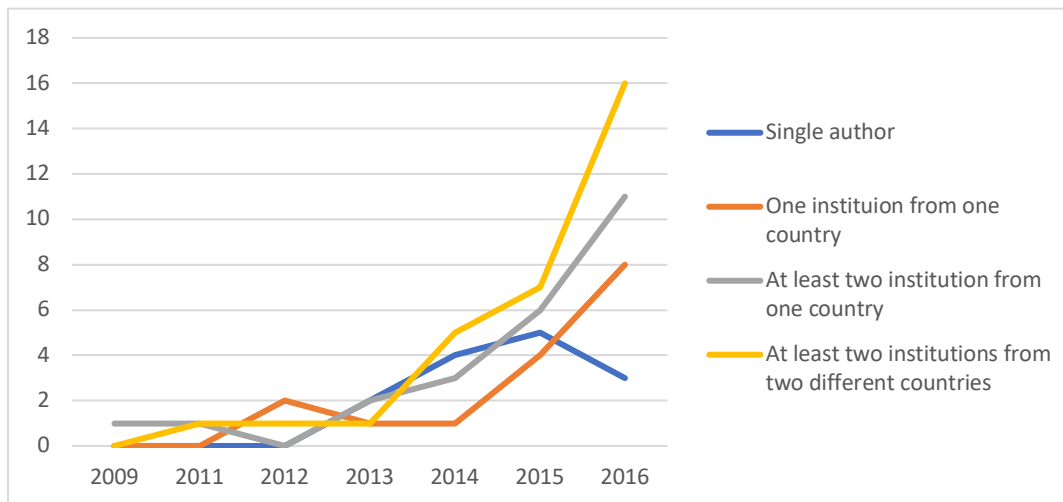


Figure 4. Collaboration Types in Big Data Articles by Year

Contributors

Author-Level. 240 unique authors contributed to the field. While 16 of them contributed with two papers, the rest (224) contribute once. The name of 16 authors are Barbara H Wixom, David Martens, Dursun Delen, Galit Shmueli, Ioanna D Constantiou, Jannis Kallinikos, John Collins, Kenneth Strang, Marijn Janssen, Ricardo Colomo Palacios, Robert J. Kauffman, Roger H. L. Chiang, Ting Li, Vladimir Stantchev, Wolfgang Ketter, and Zhaohao Sun.

Institution-Level. 136 unique institutions contributed the field. Erasmus University has the most appearance (11) in the papers, followed by Nanjing University (7), University of Alicante (7), Delft University of Technology (6), and Beihang University, Chinese Academy of Sciences, University of Minnesota with five appearances. 82 institutions of 136 institutions has one appearance in the papers.

Country- Level. Table 3 shows frequency of countries appeared in the papers.

Table 3. Contributors as Country

Countries	Frequency	Percent
USA	106	42.1
China	32	12.7
The Netherlands	20	7.9
Spain	14	5.6
Germany	13	5.2
UK	10	4.0
Italy	8	3.2
Republic of Korea	8	3.2
Singapore	6	2.4
Belgium	4	1.6
Australia	3	1.2
Austria	3	1.2
Greece	3	1.2
India	3	1.2
Liechtenstein	3	1.2
Switzerland	3	1.2
Denmark	2	.8
Norway	2	.8
Papua New Guinea	2	.8
Turkey	2	.8
Canada	1	.4
Ireland	1	.4
Israel	1	.4
Taiwan	1	.4
UAE	1	.4
Total	252	100.0

Research Themes

Table 4 summarizes the themes emphasized in the articles.

Table 4. Themes of Big Data Articles by Years

Areas	Frequency	Percent
Analytics	25	28.7
Framework/Architecture	9	10.3
Strategy	8	9.2
Supply Chain Management	6	6.9
Modeling	5	5.7
Business Intelligence	5	5.7
Governance/Policy	5	5.7
Innovation	4	4.6
Information Systems	4	4.6
Marketing	3	3.4
Knowledge Management	2	2.3
Finance	2	2.3
Text mining	2	2.3
Social Media	2	2.3
Revenue Management	1	1.1
Criminal Justice	1	1.1
Commerce	1	1.1
Geoscience	1	1.1
Entertainment	1	1.1
Total	87	100.0

Nature of Articles

Table 5 summarizes the nature of the articles.

Table 5. Nature of Big Data Articles

Nature/Study Type	Total
<i>Research/Empirical</i>	26
Theoretical or Conceptual	33
Case Study	13
Commentary	12
Applied	2
Literature Review	1
Total	87
<i>Research Methods</i>	
Quantitative	39
Not Applicable	39
Hybrid	5
Qualitative	3
Total	87

Nature of Sample

Sample as Organization.

Table 6 shows the sample types of articles based on organization, people, and country.

Table 6. Sample as Organization

Sample Type	Frequency	Percent
Not indicated	55	47.9
Transportation	7	8.0
ICT	5	5.7
Online Retail	4	4.6
Manufacturing	3	3.4
Retail	3	3.4
Public Sector	2	2.3
Consulting	2	2.3
Cultural Heritage	1	1.1
Library	1	1.1
Multinational Companies	1	1.1
E-Tourism	1	1.1
Education	1	1.1
Bank	1	1.1
Total	87	100

Sample as People.

Table 7 shows the sample types of articles based on people.

Table 7. Sample as People

Sample Type	Frequency	Percent
Not Indicated	72	82.8
Customers	7	8.0
Managers	3	3.4
Passengers	1	1.1
Investors	1	1.1
Viewers	1	1.1
Experts	1	1.1
Students	1	1.1
Total	87	100

Sample as Country.

Table 8 shows the sample types of articles based on country.

Table 8. Sample as Country

Sample Type	Frequency	Percent
Not Indicated	59	67.8
USA	12	13.8
China	5	5.7
The Netherlands	3	3.4
India	2	2.3
Europe	1	1.1
Canada	1	1.1
Belgium	1	1.1
Singapore	1	1.1
Korea	1	1.1
Germany	1	1.1
Total	87	100

LIMITATIONS AND FUTURE RESEARCH

After completing the data analysis, the researchers discovered a subtle shift change in the terminology associated the field. For example, as shown in Figure 5, the search term “machine learning” showed a significant increase in searches over time from 2014 (16) to 2019 (100). In the same period, big data changed from 51 to 62. This is just one example to highlight the changing landscape.

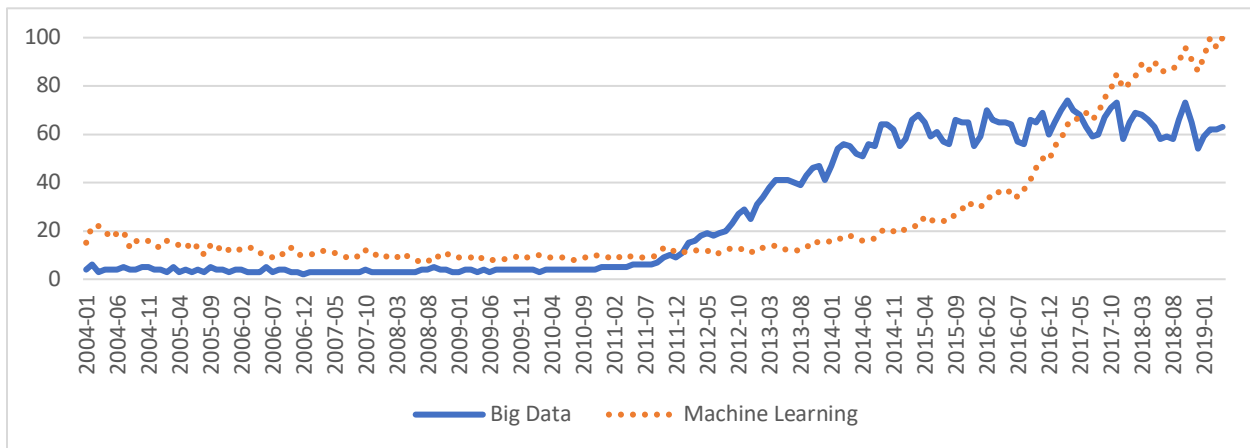


Figure 5. Worldwide Interest Over Time for Big Data (Google, 2019)

Based on this subtle shift change in the terminology associated the field, the researchers believe a more comprehensive study is necessary. Given this recognition, the data was presented with little interpretation to allow future researchers to decide the nature of additional projects. As stated in the introduction, the purpose of this project was to chronicle the progress of big data studies with a view to sparking additional research. The first phase of the new project would be to reoperationalize the term big data, perhaps in a wider context, to capture the essence of various research projects. This widening would almost certainly encompass the related field of artificial intelligence, machine learning and others. Notwithstanding the need for a broader study, we remain confident this pioneering bibliometric research will provide a foundation for future projects.

SUMMARY

This systematic analysis of big data research noted the following:

1. Most (71.2%) were published in the last two years of the review (2015 and 2016).
2. Four journals represented more than 50% of the total articles published: Decision Support Systems (19.5%), Information & Management (12.6%), Management Information System Quarterly (12.6%), and Journal of Information Technology (8.0%).
3. Most articles were written by two (28.7%) or three (27.6%) authors.
4. Most (50.6%) papers were Interdisciplinary.
5. Sixteen of the 240 unique authors participated in more than one article.
6. The most common themes were Analytics (28.7%), Framework/Architecture (10.3%), and Strategy (9.2%)
7. Authors from at least 25 countries contributed to articles. American authors lead the production with 42.1%, followed by Chinese at 12.7% and Dutch with 7.9%.

REFERENCES

- Cukier, K., & Mayer-Schoenberger, V. (2013). The Rise of Big Data. *Foreign Affairs*, 92(3), 27-40.
- Enriquez, J. (2011, October). The Glory of Big Data. *Popular Science*, 279(5), 31-33.
- Google. (2016). *Big Data Interest Over Time*. Retrieved from Google Trends:
<https://trends.google.com/trends/explore?date=2004-01-01%202016-06-01&q=big%20data>
- Koseoglu, M. A., Akdeve, E., Gedik, I., & Bertsch, A. (2015). A bibliometric analysis of strategic management articles in healthcare management literature: Past, present, and future. *International Journal of Healthcare Management*, 8(1), 27-33.
- Lowry, P. B., Moody, G. D., Gaskin, J., Galletta, D. F., Humphreys, S. L., Barlow, J. B., & Wilson, D. W. (2013). Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value? *MIS Quarterly*, 37(4), 993-A921.
- McAfee, A., & Brynjolfsson, E. (2012, October). Big Data: The Management Revolution. *Harvard Business Review*, pp. 60-68.
- Ocak, S., Koseoglu, M. A., & Yildiz, M. (2017). Business ethics research in healthcare management: A systematic review. *International Journal of Healthcare Management*, 1-7.
- Serenko, A., & Bontis, N. (2013). Global ranking of knowledge management and intellectual capital academic journals: 2013 update. *Journal of Knowledge Management*, 17(2), 307-326.
- Serenko, A., & Bontis, N. (2017). Global ranking of knowledge management and intellectual capital academic journals: 2017 update. *Journal of Knowledge Management*, 21(3), 675-692.
- Tucker, P. (2013, July-August). Mapping the future with bid data. *The Futurist*, 47(4), 14-19.