Business Process Management Journal

Big data integration with business processes: a literature review

Samuel Fosso Wamba, Deepa Mishra,

Article information:

To cite this document:
Samuel Fosso Wamba, Deepa Mishra, (2017) "Big data integration with business processes: a literature review", Business Process Management Journal, Vol. 23 Issue: 3, doi: 10.1108/BPMJ-02-2017-0047
Permanent link to this document:
http://dx.doi.org/10.1108/BPMJ-02-2017-0047

Big Data Integration with Business Processes: A Literature Review

Abstract

Purpose The purpose of this study is to improve the understanding of the integration of business process management (BPM), business process re-engineering (BPR) and business process innovation (BPI) with big data. It focuses on synthesising research published in the period 2006–16 to establish both what we know and do not know about this topic, identifying areas for future research.

Design/methodology/approach The research is based on a review of 49 published papers on big data, BPM, BPR and BPI in the top journals in the field 2006–16.

Findings In this article, we have identified the most influential works based on citations and PageRank methods. Through network analysis we identify four major clusters that provide potential opportunities for future investigation.

Practical implications It is important for practitioners to be aware of the benefits of big data, BPM, BPR and BPI integration. This paper provides valuable insights for practitioners.

Originality/value This paper is based on a comprehensive literature review, which gives big data researchers the opportunity to understand business processes in depth. In addition, highlighting many gaps in the current literature and developing an agenda for future research, will save time and effort for readers looking to research topics within big data and business processes.

Keywords: Big data analytics, research methodology, bibliometric analysis, network analysis

Paper type: Literature review

1. Introduction

The area of big data has received increasing attention from both the academic and the business communities over the past few decades. It helps to gain business insights, competitive advantage and transforms entire business processes (Wong, 2012; Oh et al., 2012; Mishra et al., 2016a). It can "create significant value for the world economy", which "enhances the productivity and competitiveness of companies and the public sector and creates a substantial economic surplus for consumers" (Manyika et al. 2011: p. 1). Several objectives can be fulfilled by analyzing big data, so there is a need for analytical techniques (big data analytics or BDA) to deal with large and complex data sets. Some organizations view BDA as a tool that can help in making strategic decisions, while scholars use it as a basis for verifying existing models and theories (Muhtaroglu et al., 2013). Organizations can empower their customers and improve decisions if they harness the power of BDA effectively (Miller, 2013). By recognizing BDA's potential, organizations improve the efficiency and quality of their business processes through effective business process management (BPM). BPM not only improves processes but also monitors the technological advances that can be integrated in the development of efficient processes through business process reengineering (BPR) and business process innovation (BPI) (Anand et al. 2013). Thus, the successful integration of BDA and business processes may create a "new class of economic asset" and help the top-performing organizations redefine their business and outperform their competitors.

A number of literature reviews on big data have been completed in the past few years (Sagiroglu and Sinanc 2013; Wamba et al., 2015; Gandomi and Haider 2015; Khorheh et al., 2015; Wang et al., 2016; Mishra et al. 2016a). Anand et al. (2013) reviewed the literature on BPM, BPR, and BPI. Nevertheless, there has been no effort to review the integration of big data and BPM, BPR and BPI, using rigorous bibliometric tools to make them useful for researchers and practitioners. There is therefore a need to synthesize the evidence about the usefulness of existing studies.

Motivated by this lack, our main objective with this study is to introduce the idea of using a bibliometric and network analysis technique to explore the world of big data and business process. We aim to: (1) systematically and rigorously collect and analyse existing studies in this field to identify the top contributing authors, countries, affiliations and key research topics; and (2) use network analysis to reveal future research gaps that can be pursued by the big data research community. We performed this analysis using the guidelines proposed by Fahimnia et al. (2015). A bibliometric and network analysis is a powerful tool for identifying established and emerging topical areas. This review collects and analyses 49 articles on big data and business processes published from January 2006 to October 2016. We believe that this review will be valuable for researchers who want to identify areas that have been thoroughly researched or where research is lacking, and for practitioners who want to stay up to date about the state of research and big data and business processes.

The paper is structured as follows. Section 2 outlines our research methodology, including protocol development, study selection, data extraction and analysis. Sections 3 and 4 report the results of the bibliometric and network analysis. Our findings, limitations and directions for future research are discussed in Section 5.

2. Methodology

This study is a bibliometric and network analysis review, that is, we document all the available studies relevant to a current area or a specific research question (Fahimnia et al., 2015). We determined on this methodology for a number of reasons: (1) to identify the top contributing authors, organisations and countries related to the field; (2) to compare citation and PageRank analysis; and (3) to uncover current research gaps through data clustering. To achieve our research objectives, we took a five-step approach, outlined in detail below: (1) develop a review protocol; (2) identify inclusion and exclusion criteria; (3) explain the search strategy process; (4) study the selection process; and (5) use data extraction and analysis.

2.1 Review protocol

Our search began with the development of a comprehensive review protocol based on the guiding principles and procedures of the bibliometric and network analysis review. This protocol identifies the search strategy, research objectives, data extraction, criteria for article selection and data analysis..

2.2 Inclusion and exclusion criteria

To achieve our objectives, we set up inclusion and exclusion criteria so that the most relevant articles were extracted from the database. We considered research articles from peer-reviewed journals in the English language, published from January 2000 to October 2016 in the Web of Science (WoS) database. We eliminated conferences, workshops, editorials, meetings, notes and tutorial summaries and considered articles only related to big data and business processes.

2.3 Search strategy

We chose the WoS database as it is one of the largest bibliographic databases, providing access to articles published since 1970 and covering approximately 8,500 high-impact research journals.

We searched for specific keywords derived from our research objectives and the structure of this review to identify relevant articles. We searched in titles, abstracts and keywords of articles in the WoS database for "business process management", "business process re-engineering", "business process innovation", "big data", "business analytics" and "big data analytics". The initial search resulted in 1,078 articles. The results were then saved in plain text format and contained basic information about the paper, such as title, authors' names and affiliations, abstracts, keywords and references.

2.4 Study selection process

Of the 1,078 studies selected, 253 were duplicates and were removed using Endnote. To fulfil the objective of our study, we restricted our search to titles, abstracts, keywords and peer-reviewed journals (excluding grey literature – workshops, conference papers, notes, editorials, meetings, etc.). This elimination process resulted in 486 relevant documents, published during the 11-year period 2006–16. The next step in the selection process was to consider articles published in the top 10 journals (i.e. the journals with the maximum papers in the field according to WoS). We found 49 articles. The distribution of the primary studies throughout the period is presented in Figure 1.

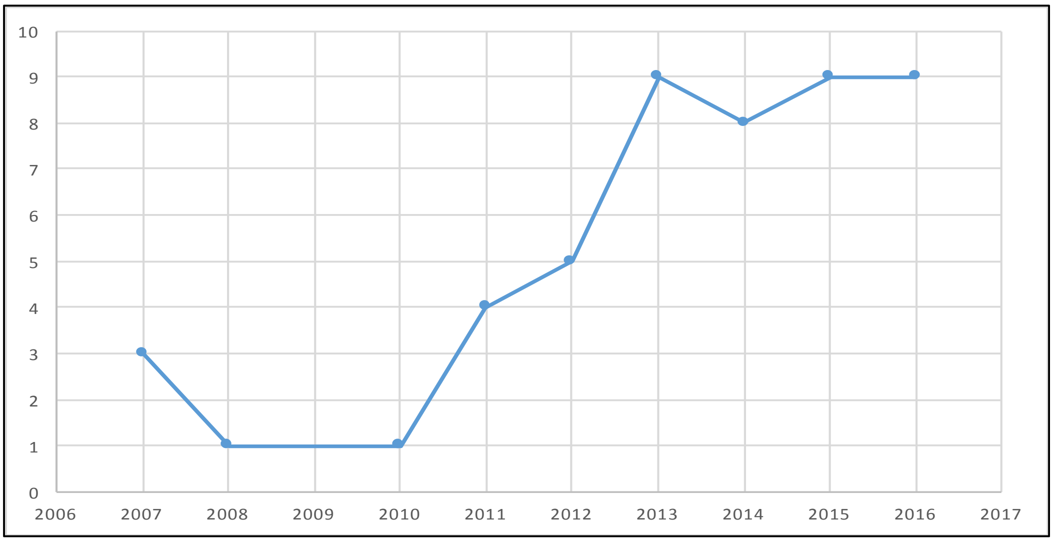

Figure 1 Yearly evolution of publications 2006–16

We restricted the period studied to 2006–16 to facilitate graphical representation. The number of papers (nine) for 2016 was estimated up to October 2016. The figure demonstrates the changing behaviour of publications in each year. It shows that the number of publications on big data and business processes increased slowly from 2006 to 2012, with a dramatic rise in publications after that date. It is clear that interest in integrating big data with business processes has increased rapidly in the past four years.

2.4.1 Distribution of articles per journal

Table 1, which details the number of articles related to big data integration with business processes by journal, shows that the most popular is *Decision Support Systems* with 13 articles (26.53%), followed some distance behind by *Interfaces* and *Industrial Management Data Systems* (5; 10.20%) and the *IBM Journal of Research and Development* (4; 8.16%).

Table 1: Number of articles per journal

Journal	No.	Percentage
Decision Support Systems	13	26.53
Interfaces	5	10.204
Industrial Management Data Systems	5	10.204
IBM Journal of Research and Development	4	8.162
Expert Systems with Applications	3	6.122
Computers in Industry	3	6.122
Software and Systems Modelling	2	4.082
Mathematical Problems in Engineering	2	4.082
Journal of the Association for Information Systems	2	4.082
International Journal of Production Research	2	4.082
IEEE Transactions on Engineering Management	2	4.082
International Journal of Information Technology and Decision Making	2	4.082
International Journal Production Economics	2	4.082
Simulation Modelling Practice and Theory	1	2.041
Wireless Personal Communications	1	2.041
Total	49	100

2.4.2 Classification of articles: approaches used most

The distribution of articles by approach used is presented in Table 2. The vast majority of papers (14; 28.57%) are experiments/model-based papers, followed by review papers (10; 20.41%) and survey studies (9; 18.37%). The results also indicate that there is a shortage of simulations (3; 6.12%) and case studies (6; 12.24%).

Table 2: Approaches used most

Approaches used most	No. of articles	Percentage
Reviews	10	20.41
Experiments/Models	14	28.57
Case studies	6	12.24
Simulation approach	3	6.12
Survey studies	9	18.37
Analytical approach	7	14.29
Total	49	100

2.4.3 Classification of articles: research areas

Table 3 shows the classification of articles based on research areas. The largest number of published articles is in computer science (35; 71.43%), followed by operations research (27; 55.10%), engineering (16; 32.65%) and business economics (7; 14.29%). The high proportion of articles in computer science is not a surprise as this research area is at the heart of the big data revolution and has provided tools to analyse massive amounts of data.

Table 3: Research areas

Research areas	Records	Percentage
Computer science	35	71.429
Operations research management science	27	55.102
Engineering	16	32.653
Business economics	7	14.286
Mathematics	2	4.082
Information science library science	2	4.082

2.5 Data extraction and synthesis

In the data extraction and synthesis stage, we read all 49 studies carefully and extracted relevant data using Endnote and Excel spreadsheets. Our main objective was to obtain full and precise content records of all the primary studies. The data related to authors, keywords, ISSN, study title, publication date, location and affiliation; cited references were extracted from the WoS core collection. Once the data from the primary studies had been extracted and recorded, we performed the analysis using BibExcel and Gephi.

3. Bibliometric analysis

In this section, we provide statistics about the 49 shortlisted articles. Specifically, we studied these articles in terms of their authors, keywords, affiliations and funding agencies. We conducted bibliometric analysis using BibExcel as it is highly flexible and capable of dealing with large volumes of data; it is also compatible with other applications such as Excel, Pajek and Gephi (Persson et al., 2009; Paloviita, 2009). An additional merit of BibExcel is that it generates data for future network analysis, which is not possible with other software like HistCite or Publish or Perish.

The data extracted from WoS in plain text format (containing all the necessary bibliographic information) was used as input into BibExcel. For the data analysis, the plain text format was reformatted to generate different file types, such as .net-file, .cit-file, .oux-file and .out-file.

3.1 Author influence

To analyze the influence of authors, we extracted the author field from the data file and recorded the frequency with which all authors appeared. Table 4 presents a list of the top 10 contributing authors and the number of publications they have authored or co-authored. As we can see Chae, with three publications, dominates the list, followed by seven others with two publications.

Table 4: Top 10 contributing authors

Authors	No. of publications
Chae, B.	3
Zhao, J.L.	2
Trkman, P.	2
Sheu, C.	2
Olson, D.	2
Mccormack, K.	2
Liberatore, M.	2
De Oliveira, M.P.V	2
Zorrilla, M.	1
Zimbrao, G.	1

3.2 Keywords

The keywords and words used in the titles of papers were extracted from WoS plain text format in BibExcel, and the frequency of their occurrence was recorded. The top 20 words used in titles and most popular keywords are presented in Tables 5 and 6. From these tables, we see a uniformity in the words used in titles and in lists of keywords. For instance, both tables include *big data*, *analytics*, *performance* and *information systems*. This demonstrates clearly that the most popular keywords are actually the search words we chose for this study.

Table 5: Top 20 words in titles

Word	Frequency	Word	Frequency
Analytics	15	Information	4
Data	10	Big	4
Business	8	Mining	3
Supply	6	System	3
Enterprise	6	Systems	3
Chain	5	Framework	3
Services	5	Process	3
Management	5	Modelling	3
Performance	4	Operational	3
Analysis	4	Impact	3

Table 6: Top 20 keywords

Word	Frequency	Word	Frequency
Management	17	Competitive advantage	3
Systems	7	Design	3
Information technology	6	Design science	3
Firm performance	6	Optimisation	3
Information systems	5	System	3
Framework	5	Knowledge	3
Models	5	Intelligence	3
Model	5	Issues	3
Integration	4	Analytics	3
Big data	4	Internet	2

3.3 Affiliation statistics

To understand the impact of affiliation on the number of publications, authors' affiliations were extracted from the WoS plain text file in BibExcel. The frequency with which these affiliations occurred was used to identify the top-performing organisations, shown in Table 7. The contribution of countries can also be identified in a similar way. Table 8 shows the top 10 countries contributing to the field of big data. A comparison of Tables 4 and 7 reveal that top universities like Kansas State University and the University of Ljubljana are represented by the top-contributing authors Chae, B. and Trkman, P. Thus, it may be concluded that the performance of one or two researchers is sufficient to improve the ranking of an institute. We also notice that the majority of work has been carried out in United States of America, followed by People's Republic of China, while only a few studies have been done in Austria and Switzerland.

Table 7: Top 10 contributing organisations

Affiliation	**No. of publications**
University of Nebraska	4
Kansas State University	4
University of Ljubljana	3
IBM Software Group	3
Villanova University	2
IBM Research Division	2
IBM Corporation	2
City University of Hong Kong	2
Vlerick Business School	1
Virgnia Polytechnic Institute	1

Table 8: Top 10 contributing countries

Countries	**No. of publications**
United States of America	24
People's republic of China	9
United Kingdom	5
Germany	4
Taiwan	3
Slovenia	3
Canada	3
Brazil	3
Austria	2
Switzerland	1

4. Network analysis

The literature records the use of various tools, such as Pajek, VOSviewer, HistCite Graph Maker, and Gephi, to perform network analysis. We chose Gephi for this study as it can handle various data formats and complex datasets and generate flexible, insightful visual aids. In Gephi, the published articles act as nodes and citations as arcs or edges. To generate graphs in Gephi, a .NET file is needed and can be created using BibExcel.

4.1 Citation analysis

Citation analysis examines the frequency with which an article is cited. The number of citations of a particular article reflects its importance in that area of research (Garfield, 1972). Thus, the importance of an article can be measured as high or low, depending on the number of citations it has received. This method helps the researchers to understand how the area of research has evolved over a period of time and which articles are the most popular (Pilkington and Meredith, 2009). Although citation analysis has been criticized, it is one of the most commonly used techniques for analyzing the literature (MacRoberts and MacRoberts 2010).

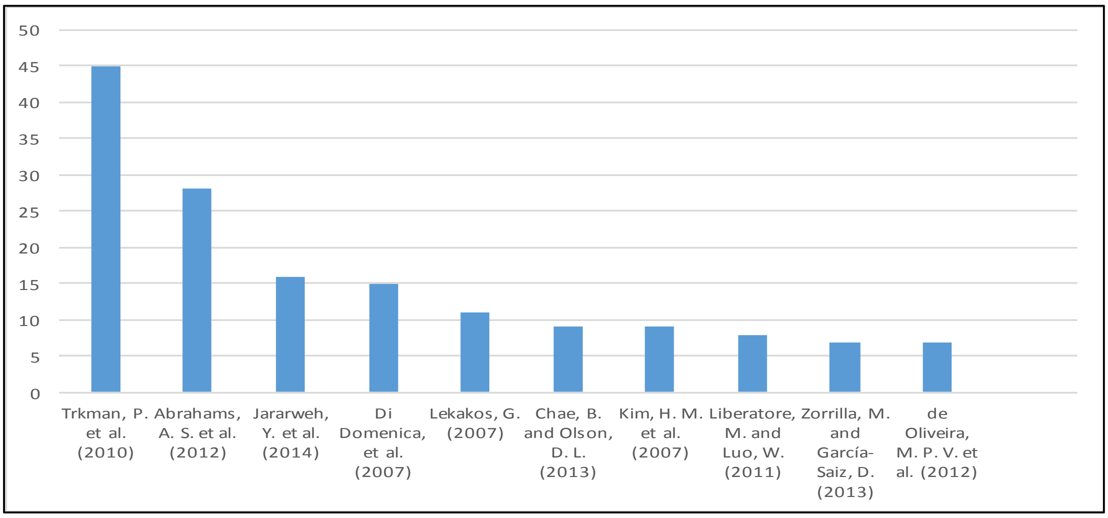

Figure 2: Top 10 cited articles: frequency

Figure 2 shows the 10 most influential works published between 2006 and 2016. The top score, 45 citations, is Trkman et al. (2010). These authors investigated the relationship between analytical capabilities and performance in the planning, sourcing, manufacturing and delivery areas of the supply chain using information system support and business process orientation as moderators. Another important contribution was made by Abrahams et al. (2012) who used a text mining technique to analyse popular online discussion forums used by motor vehicle enthusiasts. This work received 28 citations, which reflects the significance of the article in this field. The article by Jararweh et al. (2014), which has been cited 16 times, introduced a modelling and simulation environment for cloud computing known as CloudExp and integrated it with the MapReduce processing model to handle the processing of big data.

4.2 PageRank analysis

Although citation analysis is commonly used to measure the popularity of an article, Ding et al. (2009) claimed that it should not be the only criterion of an article's significance. Prestige, which records the number of times an article has been cited by other highly cited articles, is another important criterion. To account for both popularity and prestige, Brin and Page (1998) introduced PageRank, which is an excellent way to prioritise the results of web keyword searches (Mishra et al., 2016a, b). There may be situations where these two measures are positively correlated, but it is not essential for a highly cited article to be a prestigious article as well. If we compare Figure 2 and Table 9, Trkman et al. (2010) has shifted to fourth position in the list of top 10 PageRank papers, which is dominated by Lavalle et al. (2011), while none of the other articles in Figure 2 appears anywhere in Table 9.

Table 9: PageRank's top 10 articles

Article	**Page Rank**
Lavalle, S. (2011)	0.012076
Davenport, T. H. (2007)	0.011495
Chen, H. C. (2012)	0.010823
Trkman, P. (2010)	0.009516
Manyika, J. (2011)	0.008468
Bharadwaj, A.S. (2000)	0.008252
Pfeffer, J. (2006)	0.007304
Barney, J. (1991)	0.006890
Adomavicius, G. (2005)	0.006674
Feng, Y. (2008)	0.006359

4.3 Co-citation analysis

Through co-citation analysis, we can examine the relationship between groups of authors, topics, journals or keywords and explain how they are related to each other. Co-citation analysis can be based on authors or publications; author-based co-citation analysis helps depict social structure while publication-based co-citation analysis helps explain the intellectual structure of research field (Chen et al., 2010).

In this study, we used Gephi to perform co-citation analysis. When the .NET file for 49 articles is opened for the first time in Gephi, a random graph with no clear pattern is generated. To provide visibility in the graph, we used Gephi's ForceAtlas layout, in which strongly connected nodes move to the centre of the network while less connected nodes move to its boundaries (Bastian et al., 2009). This means that co-cited articles remain connected together while articles that are rarely co-cited are distanced from them. The nodes or 'outliers' isolated from the network are excluded for the purpose of data clustering, which is explained in the following section.

4.3.1 Data clustering

The data clustering method helps to group articles in different clusters (Radicchi et al., 2004; Mishra et al., 2016a, b) that has been used in literature for classifying given set of publications and also termed as modularity. The edges between the nodes in the same cluster are denser than those in different clusters. This density can be measured through modularity, an in-built tool in Gephi based on Louvain algorithm. The value of the modularity index lies in between −1 and +1 that measures the density of links inside communities versus the links between communities (Fahimnia et al., 2015). Using this algorithm, we created four major clusters and found the modularity index to be 0.49 (see Figure 3). This value indicates strong relationship between the nodes within each cluster and yet a relatively strong relationship between the nodes of different clusters.

According to Hjørland (2013), if two or more articles are cited together, they are more likely to share a similar area of interest. Therefore, we performed a detailed analysis of the papers within each cluster to identify their research area. In Table 10, we record the top publications based on their PageRank co-citation.

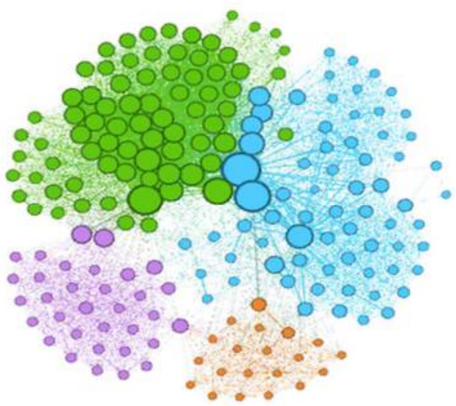

Figure 3: Four major clusters

The classification in Table 10 reveals that the articles in cluster 1 mainly focus on conceptual and theoretical studies of big data. They highlight the need for analytical tools to deal with the massive amount of data that is being generated through recent developments in technology. These works also inspired organisations to use analytical decisions for business problem-solving and competitive advantage. Motivated by the works in cluster 1, researchers in cluster 2 identify the role of business analytics in managing and solving supply chain-related problems. The majority of the articles in this cluster are empirical and focus on the techniques that help to improve supply chain performance. Cluster 3 mainly concentrates on developing methods and models that would be beneficial while dealing with forecasting problems, while researchers in cluster 4 focus on recent advances and trends in the big data environment. The first two clusters are the more popular, while there is a scope for future work in clusters 3 and 4. This four-cluster classification provides a reliable guide for scholars looking for current research topics and future research opportunities.

5. Discussion and conclusions

In this study, we present an overview of the distribution of publications on big data and business processes by conducting a bibliometric and network analysis review of articles written during the period 2006–16. To extract relevant studies, we searched for papers in the WoS database using predefined keywords. We screened papers by analysing their titles and abstracts and removed those that violated the inclusion criteria. To provide an overview of big data and business process status, we identified a primary set of 49 articles. The results of our study identify the key contributing authors, countries, affiliations and keywords across a broad spectrum of disciplines.

We can see from the bibliometric results that a large majority of the 49 primary studies were carried out in United States of America (approximately 50%), while only 2–4% were done in Switzerland and Austria. We therefore recommend that these countries should put more research effort into improving their business processes by recognising the potential of big data. Our findings also note that relatively low number of publications have appeared in this field. From Figure 1 we can see a rapid increase in publication numbers in the field of big data and business processes since 2012. This clearly demonstrates a growing interest in this area, which is unsurprising for a relatively new concept.

Reviewing and summarising what we know in relation to big data, business processes and how organizations integrate them to their advantage, we believe that this study will be beneficial for a wide range of researchers and practitioners. Our findings may help researchers to identify new research questions, gain an overview of current research and position and align their own work. Our study also helps practitioners to understand the practical challenges when integrating big data with business processes. Young scholars may use these findings as a guide to where to locate and publish different types of related research and gain further insights into the emerging field of big data.

Table 10: Top 10 papers in each cluster (PageRank co-citation measure)

Cluster 1	**Cluster 2**
Davenport and Harris (2007)	Davenport (2010)
Chen et al. (2012)	Trkman et al. (2010)
Lavalle et al. (2011)	Davis-Sramek et al. (2010)
Tversky and Kahneman (1974)	Oliva and Watson (2011)
Hevner et al. (2004)	Mithas et al. (2011)
Nunamaker et al. (1990)	Gustavsson and Wänström (2009)
Rai and Sambamurthy (2006)	Singh (2003)
Podsakoff et al. (2003)	Davenport and O'Dwyer (2011)
Sharma et al. (2014)	Turban et al. (2011)
Mackenzie et al. (2011)	Kohli and Grover (2008)
Cluster 3	**Cluster 4**
Chen-Ritzo et al. (2010)	Lee et al. (2013)
Feng et al. (2008)	Dutta and Bose (2015)
Mirzapour et al. (2011)	Morkos et al. (2012)
Nagurney (2010)	Nadadur et al. (2012)
Muhtaroglu et al. (2013)	Li et al. (2014)
Mishra et al. (2013)	Ostrosi et al. (2012)
Mortensen et al. (2008)	Powell and Snellman (2004)
Novoa and Storer (2009)	Preis et al. (2010)
Ngai et al. (2008)	Petropoulos et al. (2013)
Najafi et al. (2013)	Rai and Allada (2003)

5.1 Limitations and directions for further research

The literature review conducted in this paper has several limitations. Even though we adopted an established methodology, it could have limited the results as it focused only on articles that appeared in peer-reviewed academic journals published in English. This may have led to the exclusion of potentially relevant articles from the sample. We took care to include all past studies by consulting the WoS database but the selection process we used may have omitted some relevant research papers. Furthermore, we omitted a search for grey literature – this may provide material for further insights. Overall, this review provides a perspective on the state of big data research today.

References

Abrahams, A.S., Jiao, J., Wang, G.A. and Fan, W., 2012. Vehicle defect discovery from social media. *Decision Support Systems*, *54*(1), 87-97.

Anand, A., Wamba, S.F. and Gnanzou, D., 2013, June. A Literature Review on Business Process Management, Business Process Reengineering, and Business Process Innovation. In *Workshop on Enterprise and Organizational Modeling and Simulation* (pp. 1-23). Springer Berlin Heidelberg.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hyper textual web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple- perspective co-citation analysis. *Journal of the American Society for Information Science*, *61*(7), 1386– 1409.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229–2243.

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, *162*, 101–114.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*, 471–479.

Hjørland, B. (2013). Citation analysis: A social and dynamic approach to knowledge organization. *Information Processing Management*, *49*(6), 1313–1325.

Jararweh, Y., Jarrah, M., Kharbutli, M., Alshara, Z., Alsaleh, M. N. & Al-Ayyoub, M. 2014. CloudExp: A comprehensive cloud computing experimental framework. Simulation Modelling Practice and Theory, 49(180-192.

Amiri Khorheh, M., Moisiadis, F. and Davarzani, H., 2015. Socio-environmental performance of transportation systems. *Management of Environmental Quality: An International Journal*, *26*(6), pp.826-851.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. and Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. *MIT Sloan management review*, *52*(2), 21.

MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-sited influences. *Journal of the American Society for Information Science and Technology*, *61*(1), 1–12.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big Data: The Next Frontier for Innovation*. McKinsey Global Institute: Competition and Productivity.

Miller, Miller, G. (2013). 6 ways to use "big data" To increase operating margins by 60 %. Available at: http://upstreamcommerce.com/blog/2012/04/11/6-ways-big-data-increase-operating-margins-60-part-2 (Accessed 2nd of March, 16).

Mishra, D., Gunasekaran, A. Papadopoulos, T. and Childe, S. (2016a). Big Data and Supply Chain Management: A Review and Bibliometric Analysis. *Annals of Operations Research*, DOI 10.1007/s10479-016-2236-y.

Mishra, D., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Dubey, R. and Wamba, S. (2016 b). Vision, applications and future challenges of Internet of Things. *Industrial Management & Data Systems*, 116(7), 1331-1355.

Muhtaroglu, F.C.P., Demir, S., Obali, M., Girgin, C., 2013. Business model canvas perspective on big data applications. In: Proceedings of the IEEE International Conference on Big Data, pp. 32–36.

Oh, L., Teo, H., & Sambamurthy, V. (2012). The effects of retail channel integration through the use of information technologies on firm performance. *Journal of Operations Management*, *30*(1), 368–381.

Paloviita, A. (2009). Stakeholder perceptions of alternative food entrepreneurs. *Management and Sustainable Development, World Review of Entrepreneurship*, *5*(4), 395–406.

Persson, O., Danell, R., & Schneider, J. W. (2009). *How to use bibexcel for various types of bibliometric analysis*. In: In F. Åström, R. Danell, B. Larsen, & J. W. Schneider (Eds.), Celebrating scholarly com- munication studies: A Festschrift for Olle Persson at his 60th Birthday. 05-S, *International Society for Scientometrics and Informetrics*, 9–24.

Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management-1980-2006: A citation/co-citation analysis. *Journal of Operations Management*, *27*(3), 185–202.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., & Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 2658–2663.

Sagiroglu, S., & Sinanc, D. (2013). Big Data: A review. In *International conference on collaboration technologies and systems (CTS)* (pp. 42–47).

Trkman, P., McCormack, K., Oliveira, M., & Ladeira, M. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318–327.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'Big Data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246.

Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98–110.

Wong, D. (2012). *Data is the next frontier, analytics the new tool: Five trends in big data and analytics, and their implications for innovation and organisations*. London: Big Innovation Centre.

Appendix 1

Abbasi, A., Sarker, S. & Chiang, R. H. L. 2016. Big Data Research in Information Systems: Toward an Inclusive Research Agenda. Journal of the Association for Information Systems, 17(2), pp I-XXXII.

Abrahams, A. S., Jiao, J., Wang, G. A. & Fan, W. G. 2012. Vehicle defect discovery from social media. Decision Support Systems, 54(1), pp 87-97.

Afshari, H. & Peng, Q. J. 2015. Modeling and quantifying uncertainty in the product design phase for effects of user preference changes. Industrial Management & Data Systems, 115(9), pp 1637-1665.

Apte, C., Dietrich, B. & Fleming, M. 2012. Business leadership through analytics. Ibm Journal of Research and Development, 56(6), pp 5.

Babiceanu, R. F. & Seker, R. 2016. Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. Computers in Industry, 81(128-137.

Barenfanger, R., Otto, B. & Osterle, H. 2014. Business value of in-memory technology - multiple-case study insights. Industrial Management & Data Systems, 114(9), pp 1396-1414.

Basole, R. C., Seuss, C. D. & Rouse, W. B. 2013. IT innovation adoption by enterprises: Knowledge discovery through text analytics. Decision Support Systems, 54(2), pp 1044-1054.

Butler, C. A. & Robinson, R. S. 2012. Introduction: 2011 Daniel H. Wagner Prize for Excellence in Operations Research Practice. Interfaces, 42(5), pp 421-424.

Cao, G. M., Duan, Y. Q. & Li, G. D. 2015. Linking Business Analytics to Decision Making Effectiveness: A Path Model Analysis. Ieee Transactions on Engineering Management, 62(3), pp 384-395.

Caron, F., Vanthienen, J. & Baesens, B. 2013. A comprehensive investigation of the applicability of process mining techniques for enterprise risk management. Computers in Industry, 64(4), pp 464-475.

Chae, B. 2014. A complexity theory approach to IT-enabled services (IESs) and service innovation: Business analytics as an illustration of IES. Decision Support Systems, 57(1-10.

Chae, B., Olson, D. & Sheu, C. 2014a. The impact of supply chain analytics on operational performance: a resource-based view. International Journal of Production Research, 52(16), pp 4695-4710.

Chae, B., Yang, C. L., Olson, D. & Sheu, C. 2014b. The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. Decision Support Systems, 59(119-126.

Chae, B. K. & Olson, D. L. 2013. Business analytics for supply chain: a dynamic-capabilities framework. International Journal of Information Technology & Decision Making, 12(1), pp 9-26.

de Oliveira, M. P. V., McCormack, K. & Trkman, P. 2012. Business analytics in supply chains - The contingent effect of business process maturity. Expert Systems with Applications, 39(5), pp 5488-5498.

Di Domenica, N., Mitra, G., Valente, P. & Birbilis, G. 2007. Stochastic programming and scenario generation within a simulation framework: An information systems perspective. Decision Support Systems, 42(4), pp 2197-2218.

Dutta, D. & Bose, I. 2015. Managing a Big Data project: The case of Ramco Cements Limited. International Journal of Production Economics, 165(293-306.

Fabian, B., Kunz, S., Muller, S. & Gunther, O. 2013. Secure federation of semantic information services. Decision Support Systems, 55(1), pp 385-398.

Farwick, M., Schweda, C. M., Breu, R. & Hanschke, I. 2016. A situational method for semi-automated Enterprise Architecture Documentation. Software and Systems Modeling, 15(2), pp 397-426.

Hampapur, A., Cao, H., Davenport, A., Dong, W. S., Fenhagen, D., Feris, R. S., Goldszmidt, G., Jiang, Z. B., Kalagnanam, J., Kumar, T., Li, H., Liu, X., Mahatma, S., Pankanti, S., Pelleg, D., Sun, W., Taylor, M., Tian, C. H., Wasserkrug, S., Xie, L., Lodhi, M., Kiely, C., Butturff, K. & Desjardins, L. 2011. Analytics-driven asset management. Ibm Journal of Research and Development, 55(1-2), pp 19.

Hargaden, V. & Ryan, J. K. 2015. Resource Planning in Engineering Services Firms. Ieee Transactions on Engineering Management, 62(4), pp 578-590.

Hwang, J., Bai, K., Tacci, M., Vukovic, M. & Anerousis, N. 2016. Automation and orchestration framework for large-scale enterprise cloud migration. Ibm Journal of Research and Development, 60(2-3), pp 12.

Jararweh, Y., Jarrah, M., Kharbutli, M., Alshara, Z., Alsaleh, M. N. & Al-Ayyoub, M. 2014. CloudExp: A comprehensive cloud computing experimental framework. Simulation Modelling Practice and Theory, 49(180-192.

Kim, H. M., Fox, M. S. & Sengupta, A. 2007. How to build enterprise data models to achieve compliance to standards or regulatory requirements (and share data). Journal of the Association for Information Systems, 8(2), pp 105-128.

Kowalczyk, M. & Buxmann, P. 2015. An ambidextrous perspective on business intelligence and analytics support in decision processes: Insights from a multiple case study. Decision Support Systems, 80(1-13.

Ku, E. C. S. 2014. Functional integration and systems implementation of customer relationship management in hotel industry: a multilevel analysis. International Journal of Information Technology & Decision Making, 13(1), pp 175-196.

Lan, S. L., Zhang, H., Zhong, R. Y. & Huang, G. Q. 2016. A customer satisfaction evaluation model for logistics services using fuzzy analytic hierarchy process. Industrial Management & Data Systems, 116(5), pp 1024-1042.

Lech, P. 2012. Information gathering during enterprise system selection: insight from practice. Industrial Management & Data Systems, 112(5-6), pp 964-981.

Leitner, A., Preschern, C. & Kreiner, C. 2014. Effective development of automation systems through domain-specific modeling in a small enterprise context. Software and Systems Modeling, 13(1), pp 35-54.

Lekakos, G. 2007. Exploiting RFID digital information in enterprise collaboration. Industrial Management & Data Systems, 107(8), pp 1110-1122.

Li, Q., Luo, H., Xie, P. X., Feng, X. Q. & Du, R. Y. 2015. Product whole life-cycle and omni-channels data convergence oriented enterprise networks integration in a sensing environment. Computers in Industry, 70(23-45.

Liberatore, M. & Luo, W. 2013. ASP, The Art and Science of Practice: A Comparison of Technical and Soft Skill Requirements for Analytics and OR Professionals. Interfaces, 43(2), pp 194-197.

Liberatore, M. & Luo, W. H. 2011. INFORMS and the Analytics Movement: The View of the Membership. Interfaces, 41(6), pp 578-589.

Olavson, T. & Fry, C. 2008. Spreadsheet decision-support tools: Lessons learned at Hewlett-Packard. Interfaces, 38(4), pp 300-310.

Park, S. T., Kim, Y. R., Jeong, S. P., Hong, C. I. & Kang, T. G. 2016. A Case Study on Effective Technique of Distributed Data Storage for Big Data Processing in the Wireless Internet Environment. Wireless Personal Communications, 86(1), pp 239-253.

Peiris, K. D. A., Jung, J. & Gallupe, R. B. 2015. Building and evaluating ESET: A tool for assessing the support given by an enterprise system to supply chain management. Decision Support Systems, 77(41-54.

Perko, I., Gradisar, M. & Bobek, S. 2011. Evaluating probability of default: Intelligent agents in managing a multi-model system. Expert Systems with Applications, 38(5), pp 5336-5345.

Silva, L. A. E., Siqueira, M. F., Pinto, F. D., Banos, F. S. M., Zimbrao, G. & Souza, J. M. 2016. Applying data mining techniques for spatial distribution analysis of plant species co-occurrences. Expert Systems with Applications, 43(250-260.

Song, X., Wu, Y. L., Ma, Y. F., Cui, Y. & Gong, G. H. 2015. Military Simulation Big Data: Background, State of the Art, and Challenges. Mathematical Problems in Engineering, 20.

Spaulding, T. J., Furukawa, M. F., Raghu, T. S. & Vinze, A. 2013. Event sequence modeling of IT adoption in healthcare. Decision Support Systems, 55(2), pp 428-437.

Sun, S. X. & Zhao, J. L. 2013. Formal workflow design analytics using data flow modeling. Decision Support Systems, 55(1), pp 270-283.

Trkman, P., McCormack, K., de Oliveira, M. P. V. & Ladeira, M. B. 2010. The impact of business analytics on supply chain performance. Decision Support Systems, 49(3), pp 318-327.

Wang, C. H. & Chien, Y. W. 2016. Combining balanced scorecard with data envelopment analysis to conduct performance diagnosis for Taiwanese LED manufacturers. International Journal of Production Research, 54(17), pp 5169-5181.

Wang, G., Gunasekaran, A., Ngai, E. W. T. & Papadopoulos, T. 2016. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, 176(98-110.

Wang, H. J. & Zhao, J. L. 2011. Constraint-centric workflow change analytics. Decision Support Systems, 51(3), pp 562-575.

Yesudas, M., Menon, G. & Ramamurthy, V. 2014. Intelligent operational dashboards for smarter commerce using big data. Ibm Journal of Research and Development, 58(5-6), pp 10.

Yu, J. F. & Hoff, R. 2013. Optimal Routing and Assignment of Consultants for Energy Education, Inc. Interfaces, 43(2), pp 142-151.

Zhao, D. F., Zhao, W., Sun, L. & Huang, D. M. 2015. Modeling and Querying Business Data with Artifact Lifecycle. Mathematical Problems in Engineering, 9.

Zorrilla, M. & Garcia-Saiz, D. 2013. A service oriented architecture to provide data mining services for non-expert data miners. Decision Support Systems, 55(1), pp 399-411.