Arvidsson, Ala Pazirandeh; Jonsson, Patrik; Kaipia, Riikka

Big data in purchasing and supply management

Published in:
International Journal of Procurement Management

DOI:
10.1504/IJPM.2021.113490

Published: 01/01/2021

Document Version
Peer reviewed version

Please cite the original version:
Arvidsson, A. P., Jonsson, P., & Kaipia, R. (2021). Big data in purchasing and supply management: A research agenda. *International Journal of Procurement Management*, *14*(2), 185-212.
https://doi.org/10.1504/IJPM.2021.113490

Forthcoming in the: International Journal of Procurement Management **ISSN online** 1753-8440 **ISSN print** 1753-8432

Big data in purchasing and supply management: A research agenda

Ala Pazirandeh Arvidsson*

Patrik Jonsson*

Riikka Kaipia*, ** (riikka.kaipia@aalto.fi)

* Department of Technology Management and Economics, Chalmers University of Technology, Gothneburg, Sweden, SE-412 96

** Department of Industrial Engineering and Management, Aalto University, Aalto, Finland, Box 15500, 00076

Dr. Ala Pazirandeh is an associate professor at Chalmers university and has a PhD in industrial management and logistics engineering from Lund University. Her research is related to the supply chain collaborations with a large focus on innovation, sustainability, and digitalization. She has insight from the industrial, public and humanitarian sector. She is the author and reviewer of several peer-reviewed scientific publications in scientific purchasing, logistics, and operations management journals.

Patrik Jonsson is professor in Operations & Supply Chain Management and the Head of Division at Supply and Operations Management. His research, teaching and outreach activities are in the field of operations and supply chain planning, with a specific interest in the generation of demand and supply coordination in firms and supply chains through data- and technology-enabled operations and planning processes. Patrik has a PhD in production management from Lund University. He is the author and reviewer of many scientific articles and serves as a regional, associate, or board member editor for several high ranked operations SCM journals.

D.Sc. (Tech.) Riikka Kaipia works as a Research Fellow at Aalto University, Finland, and as a Senior Lecturer at Chalmers University of Technology, Sweden. Her research interests are connected to purchasing and supply management, and information and coordination in supply chains, and sales and operations planning. She has published articles in high quality journals, including Journal of Operations Management and Research Policy, and received publication awards.

Forthcoming in the: International Journal of Procurement Management **ISSN online** 1753-8440 **ISSN print** 1753-8432

Abstract

This study aims to synthetize knowledge about big data (BD) in the area of purchasing and supply management (PSM) and to develop an agenda for future research. A systematic literature review approach was employed, and 65 articles were synthesized based on a classification for BD in PSM to identify research gaps and opportunities. Literature in this area is descriptive and conceptual in nature, and modeling methods are favored. BD applications were identified for supplier selection, supply risk management, and supply visibility. Based on the complexity of access and use of BD, we identify a need for contextualized theory-building studies.

Paper type: Literature review

Keywords: Big data; literature review; purchasing and supply management; data analytics; research agenda

1. Introduction

Big data (BD) offer organizations a powerful means to obtain value from the increasing amounts of data created internally and externally (Sanders, 2014; Schoenherr and Speier-Pero, 2015), characterized by volume, variety, velocity, and value (Chen, et al., 2012). BD analytics can be defined as technologies (database and data-mining tools) and techniques (analytical methods) that can be employed to analyze large-scale, complex data to affect firm performance (Kwon et al., 2014). With this definition, high-tech data storage, management, analytics capability, and visual technologies are all part of big data analytics (Chen et al., 2012; Jagadish et al., 2014). An essential challenge for companies, however, is how to generate value from BD (Stentoft et al., 2016).

Purchasing and supply management (PSM) – including activities from strategizing for supply, the selection of suppliers, the management of supplier relationships, and ordering and expedition – has been claimed to have huge potential in benefiting from big data (BD) (Moretto et al., 2017; Nguyen et al., 2017; Lamba and Singh, 2017). In this area, knowledge generation has been suggested to happen by, for example, using BD to evaluate different supply risk scenarios (Lamba and Singh, 2017; Nguyen et al., 2017), improve data-driven projections of demand in order to plan purchases (Sanders, 2014) or to better align internal and external supplier goals and processes in order to improve supply flows (Wang et al., 2016, Srai & Lorentz, 2019).

While the future of most business seems to be entangled in the competitive incorporation of BD into different functional areas including PSM (Connaughton and Sawchuk, 2014), both research and practice in the application of BD in PSM are lagging behind (Schoenherr and Speier-Pero, 2015). Very few mature application cases can be identified in practice. Based on their survey study of Danish SCM professionals, Brinch et al., (2018) contend that PSM might be a less important area for the application of BD. However, other studies suggest the opposite (Sanders, 2016; Srai & Lorentz, 2019). Such contradictory findings can be related to the lack of application solutions developed for PSM decision making. In their recently published study, Srai and Lorentz (2019) find only one published study on BD in the PSM area (Ivanov, 2017), but relatively more on cognitive computing (e.g. for intelligent supplier segmentation and selection). Kamann et al., (2018) contend BD to be an enabler of systems thinking in the PSM area. Published studies on narrower aspects of PSM are scattered in differing outlets with differing terminology, making it difficult to identify and access research on this topic, to assess the state of the research, and to develop further studies.

In order to contribute to the research on BD in PSM, new research needs to build on the extant knowledge. Thus, to identify relevant existing contributions in the literature, in this paper, we map current understanding of the application areas of BD in PSM through a systematic literature review. We aim to advance the field by studying how BD literature has addressed PSM and how research could proceed to support purchasing managers in their attempts to benefit from BD. We analyzed and synthetized data from a sample of 65 papers, and we report the evidence on what is and is not known about BD in the area of PSM. The following research questions were formulated to guide our literature review:

RQ1: How does extant research on BD address PSM-related decision-making?

RQ2: What are the needs and opportunities for future research on BD in the area of PSM?

Through the literature review, we identify and analyze 65 research papers published on the topic. We synthesize the observations from our literature review and present a classification on PSM-related BD. Our study identifies further research needs and introduces potential research questions and avenues for the development of PSM-related BD research.

This paper starts by presenting the method used and describing the process of identifying relevant literature. Next, it presents results on how the research has developed, and which methods, theories, and data are used in the identified papers. Then, summaries of BD sources, application areas, and solutions for BD in PSM are presented. Finally, the paper develops a discussion on the observations and suggests areas for further research.

2. Method

In order to generate knowledge on BD in PSM, research needs to build on the extant knowledge. Thus, a systematic process of scanning the academic literature suggested by Tranfield et al., (2003) was employed to map the state of knowledge on BD in the PSM area. A systematic literature review is used to identify the conceptual content of the domain under review, which may contribute to theory development. We followed the process for conducting a systematic literature review suitable for the plurality of management research (Tranfield et al., 2003; Rousseau et al., 2008), starting with the formulation of questions in order to clarify the purpose and the audience of the review, then moving on to a comprehensive identification of relevant research, organizing, interpreting, and finally synthesizing the content.

2.1. Identification of relevant research

A systematic selection process was applied (see Table 1). A keyword search was conducted in Google Scholar (for article titles), and Scopus and Web of Science (for abstracts and keywords). Two groups of keywords were defined based on an explorative search of some key articles, one indicating the content and various aspects of PSM (i.e., procurement, purchasing, supply, sourcing, supplier, negotiation) and another covering BD (i.e., big data, data analytics, data sciences, predictive analytics, data-based decision-making, intelligence). One keyword from both groups was required to appear in the articles (the logical operator “AND” was used). The three authors

jointly decided on the initial cleaning criteria and focus, which were to include all the articles that were in English and contained in a journal, conference, or professional outlet. One of the researchers went through the hits and, based on these criteria and after deleting double hits, ended up with 214 initial articles (by late fall 2017).

Next, the articles were selected in two review rounds, applying defined criteria for inclusion and exclusion. Articles were to be included if they provided findings connected to PSM or the upstream flow of a supply chain, and if they handled BD as a major topic. Articles handling consumer markets, or the purchasing of a BD system or BD services, were excluded. Exclusion and inclusion decisions were made based on reading abstracts of the articles, and two researchers individually color-coded each article into a database so that red indicated *excluded*, yellow *uncertain*, and green *included*. This coding ended up with uncertainty about the inclusion/exclusion of about 20–25 articles for each pair (i.e., an article was coded by two researchers as red–green, red–yellow, yellow–green, yellow–yellow), the abstracts of which were then re-read and discussed until consensus was reached. This phase ended up in a sample of 76 articles. In the second round, the full texts of the articles were read, and the same inclusion/exclusion criteria were applied. If an article did not meet the criteria, it was evaluated by another researcher and discussed in the group until a decision was made.

Next, studying the articles identified in earlier literature reviews on BD resulted in three new articles meeting the inclusion criteria. Additionally, a new keyword search conducted early in 2019 identified nine articles from the years 2017-2018 meeting the inclusion criteria. This comprehensive process resulted in a sample of 65 articles.

1.1.Organizing and interpreting

For data extraction, a systematic coding process was used (Rousseau et al., 2008). We first developed codes jointly among all the authors, based on the exploratory round and previous literature. Codes were defined to include citation information (publication year, journal, authors, title), method and data used, terms used for analytics and BD, main findings of the article, and different PSM decisions or areas affected by BD, which were divided into strategic, tactical, and operational parts of the purchasing process (adapted from Van Weele, 2010). Additional codes were allowed to emerge. An initial coding round for four articles was conducted by two researchers to test the process and codes, based on which some adjustments were made.

The articles were divided between the three authors for coding. To reduce mistakes and avoid the omission of relevant material, the researchers met frequently during the coding process to discuss uncertainties and coding procedures. The method for coding the articles was based on the description of the methodology, and an article was classed as conceptual if it was descriptive in nature and no method was explained or research data used. In this phase, initial themes and topics of analysis were identified.

Table 1 The stages of the article selection process.

Stages	**Details and sequence of activities**	**No. of articles**

Search rounds	A search with the Google Scholar, Scopus and Web of Science (fall 2017), using all combinations of the keywords from groups A and B: A_i: Procurement / purchasing / supply / sourcing / supplier / negotiation B_n: Big data / data analytics / data sciences / predictive analytics / data-based decision-making / intelligence	Hits for each keyword combination A_1–B_n 19/8/1/0/0/40 A_2–B_n 4/0/0/1/0/37 A_3–B_n 204/66/6/35/0/331 A_4–B_n 16/3/1/1/0/23 A_5–B_n 3/2/0/0/0/34 A_6–B_n 3/3/0/0/0/115
Initial cleaning	Full academic, conference and practitioner articles included, double hits and books excluded.	214 articles
Selection based on article abstracts	A selection of the articles by at least 2 researchers. - Inclusion criteria: • Connected to PSM or the upstream flow of supply chain • The article handles BD (indicated by *data analytics, data science* or a similar term) - Exclusion criteria: • Related to / planning for / purchase by the consumer market • The article handles the purchasing of BD/BD • If there are differences in opinion, the article was discussed until consensus was reached	76 articles
Selection based on the articles' full texts	Selection of the articles by one researcher using the same criteria, and based on answering *yes* to both of the following questions: 1) Are there findings related to PSM? 2) Is BD a central topic? If the answer to either question is *no*, another researcher read the article, and decision is made based on discussion if there is a difference in opinion.	49 articles
Articles from other literature reviews and recently published articles are added	Articles were identified from the reviewed articles' references and complied with our inclusion and exclusion criteria. A new keyword searches up to 2019. A review of all papers and issues in the journal of purchasing and supply management from 2014-2019	16 articles
Final sample		65 articles

1.2.Forming synthesis

All the codes and article details were collected in a single database. In the coding process, as well as during reading, themes, issues, gaps, definitions, and possible classifications surfaced that were further analyzed. The results were summarized in a number of graphs and tables, and reviewed and further developed in a team of all the authors. This required going back to the articles several times to check that the coding was correct and that all the relevant findings were captured. The results of this process are presented in the next section of the paper. Based on the observations from the literature on the important aspects of BD in PSM, we introduce a classification for this kind of data and further synthesize the findings from the literature based on the four categories to identify research gaps and opportunities. These areas are then further synthesized and discussed based on existing studies and knowledge of the PSM areas (e.g., supply risk management or supplier selection) to develop a research agenda for the field.

2. Results from the systematic literature review

Most of the articles in the final sample of 65 were from academic journals (44), the top journal being *Journal of Computers & Operations Research* with four articles. *The International Journal of Logistics Management* and *Technological Forecasting and Social Change* followed with three articles each. In terms of the disciplines of the journals, one focused on the specific area of PSM (*Journal of Purchasing and Supply Management)* and had published two papers on the topic, and four journals on logistics and SCM (*The International Journal of Logistics Management; International Journal of Logistics, Research and Application; International Journal of Physical Distribution and Logistics Management;* and *Journal of Business Logistics*). Two journals were from the area of operations management (*International Journal of Production Economics* and *International Journal of Operations and Production Management*). There were 22 articles from conference proceedings.

2.1. Development of the research over the years

The oldest articles in our sample were published in 2008. As shown in Figure 1, the topic has only attracted attention in recent years, gaining momentum from 2014 onwards, with 7, 6, 14, 17, and 14 publications per year, respectively, during 2014–2018.

Fewer than 30% of the articles (21) discussed BD in the context of PSM as their main topic (Figure 1). The articles relating BD to supply chain management dominate the sample with 30 articles. The remaining topics with findings related to the application of BD in PSM were on supply chain risk management, supply market intelligence (SMI), new product development, and e-commerce (with three and two articles for the first two topics respectively, and one article each for the last two topics).

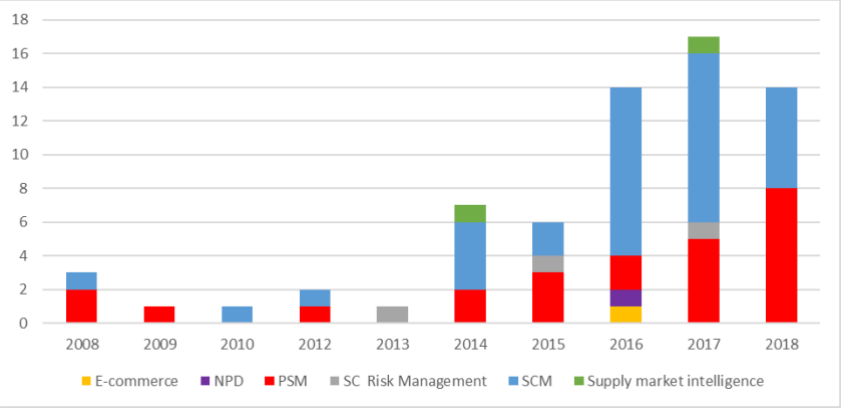


Figure 1 Identified BD articles and the context (main topic) in which BD is discussed (N=65).

2.2. Methodology used in the articles

The majority of the articles, 22, were conceptual (Figure 2), followed by 15 modeling (both analytical and simulation) articles. Altogether, eight articles were literature reviews, and eight applied multiple methods. Of these multi-method articles, two used focus groups and interviews, two applied modeling and conceptual development, one used modeling and a single case to illustrate its application, and one used modeling and empirical experiments for validation. In one

study, the Delphi method and a questionnaire survey were applied, and two articles used a combination of surveys and interviews, of which one also included a multiple case study. Altogether, there were 20 empirical papers (see Table 2), of which five applied a multi-method, five were surveys, three applied a Delphi method, four were case study papers, and two were modeling papers that used data to validate their models. Finally, one article used opinion mining as a method.

Table 2 An overview of the methods used in the empirical studies (N = 20).

Method	**Data used**	**Reference**
Multi-method	23 experts in Delphi study, 49 responses in a survey	Brinch et al., 2018
	9 interviews and a survey of 89 companies	Handfield, 2014
	4 focus groups with 15 PSM experts + an unspecified number of interviews in the cases	Moretto et al., 2017
	2 focus groups with 75 purchasing professionals + 6 semi-structured interviews	Paajanen et al., 2017
	52 open-ended interviews + 102 survey responses + 4 cases	Sanders, 2016
Survey	A survey with 481 responses	Banerjee and Mishra, 2016
	A survey with 235 suppliers using 21 criteria	Matta and Tayal, 2017
	A survey with 531 supply chain experts (6.2% purchasing)	Schoenherr and Speier-Pero, 2015
	A survey with 210 responses	Lai et al., 2018
	A survey of 287 SCM and IT professionals	Raman et al., 2018
The Delphi method	A Delphi study with 24 individuals	Brinch et al., 2017
	A Delphi study with 20 individuals	Kache and Seuring, 2017
	A Delphi study with 73 panellists	Roßmann et al., 2018
Case study	A case study: 1 manufacturer, 18 suppliers, 5 criteria	Deswal and Garg, 2015
	The PSM transactions of one company between 2011 and 2014	Tan and Lee, 2015
	27 interviews in 27 companies in 6 countries	Richey et al., 2016
	Practices from 4 different cases on digitalization with 3 having some applications in BD	Srai and Lorentz, 2019
Modelling	Netflix open sourced data - 685,000 rows of data equalling around 4,700 Users and 177 Items	Pouraghabagher and Sarfaraz, 2018
	Open source public sector data	Choi et al., 2018
Opinion mining	Unstructured social media data from 600 companies	Swain and Cao, 2017

Only a few articles in our sample are connected to higher-level theories. Liang et al., (2016) applied social network theory to the context of PSM fraud detection and by combining BD on the intimacy between buyers and bidders, abnormal interactions, commodity prices, and supplier credits. Roßmann et al., (2018) applied information processing theory to explain the application of BD as a means of increasing information processing capacity and explaining the benefits of it for managing increasing uncertainties and creating information processing requirements. Richey et al., (2016) compare the explanatory power of the dynamic capability approach and the resource-based view to explain the long-term value of BD to supply chain partnerships and conclude that many of

the characteristics of BD conform to those of dynamic capability. Srai and Lorentz (2019) build a framework for PSM-related digitalization of which BD is one aspect, based on the value-adding principles for PSM extracted from five different theories of resource-based view, transaction cost economics, agency theory, knowledge-based view, and information processing view.

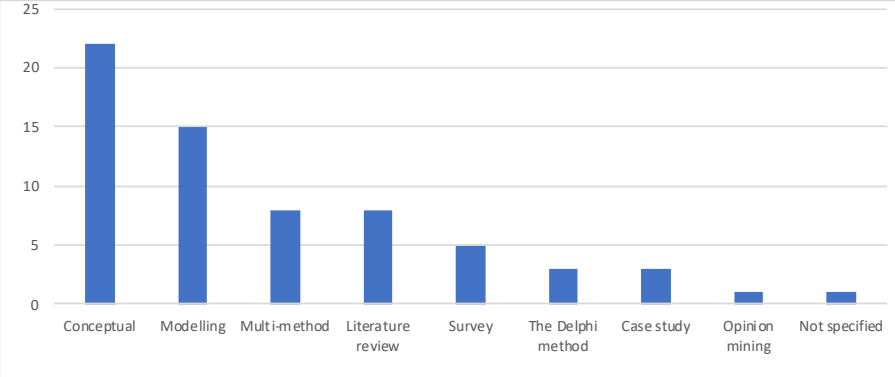


Figure 2 Distribution of the articles based on the method employed (N=65).

2.3. A closer look at the literature review papers

There are eight literature review articles in our sample, but none focuses solely on PSM as an area in which to apply BD. The majority focuses on the SCM area with relevance to PSM, with an additional article on e-commerce. The details of each literature review are summarized in Table 3. These articles report small numbers of published articles on BD in PSM-related domains. Rozados and Tjahjono (2014), Barbosa et al., (2018), and Chehbi-Gamoura and Derrouiche (2017) did not analyze the content of the publications, while Wang et al., (2016), Lamba and Singh (2017), and Nguyen et al., (2017) studied BD applications in supplier selection, supply spending/sourcing costs, and supply risk management processes. Nguyen et al., (2017) relate BD to levels of analytics and report that BD in the area of PSM is applied using descriptive, predictive, and prescriptive analytics, and Rozados and Tjahjono (2014) list 25 data sources of relevance to PSM. These reviews, consequently, identify the application of BD to PSM as an area with a small but increasing number of publications, mainly related to three broad areas of application. The literature reviews emphasize the need for BD research related to PSM.

Table 3 An overview of the literature review papers (N = 8).

Title	Author(s)	No. of articles in the review	Findings related to PSM
Bid data analytics in E-commerce	Akter and Wamba, 2016	48	BD is a distinctive competence of the high-performance business process that is conducted to support business needs, such as determining the optimal price, detecting quality, improving multi-channel integration and coordination, and enhancing global sourcing.
Understanding BD analytics capabilities in supply chain management	Arunachalam, et al., 2018	82	A maturity model of supply chain capabilities is presented that identifies organizational and technical challenges for BD. Directions for further research are suggested.

Managing supply chain resources with big data analytics	Barbosa et al., 2018	44	A general conclusion that BD is only a powerful resource when supported by capabilities (especially those related to human competences and skills) is presented.
Big valuable data in supply chain: Deep analysis of current trends and coming potential	Chehbi-Gamoura and Derrouiche, 2017	Not specified	Sourcing is identified as one of two SCOR processes (the other being making/manufacturing) that need more attention in future research. The need for data to be available across functions and processes is emphasized.
Big data in operations and supply chain management: Current trends and future perspectives	Lamba and Singh, 2017	Not specified	Future perspectives are presented on using high variety, velocity and volume data in total purchasing cost optimization and in joint problems that optimize supplier selection and carrier selection, while also optimizing dynamic lot sizes across entire planning horizons.
Big data analytics in supply chain management: A state-of-the-art literature review	Nguyen et al., 2017	88	Studies on BD in PSM are spread across topics. Proposed future research directions with relevance for PSM include the further investigation of applications to specific supply chain functions, functional alignment strategy for the horizontal integration of BD in supply chains, balancing the focus on all three analytical levels and combining analytical techniques.
Big data analytics in SCM: Trends and related research	Rozados and Tjahjono, 2014	87	PSM's need to use other BD sources than spend data is emphasized. The article lists examples of data sources and specifies 25 different data sources of relevance for PSM. The article proposes that most organizations need to make disparate data sources accessible across functions and exemplifies how BD could support the supplier negotiation process.
Big data analytics in logistics and supply chain management: Certain investigations for research and applications	Wang et al., 2016	101	The article lists spend, supplier performance assessments and negotiations as BD sources in PSM and mentions managing supplier risks and managing supplier performance as application areas. The need for more research on descriptive analytics and on BD analytics as an integrated business capability in an organization is proposed.

2.4. Findings of the empirical papers

The findings of the empirical articles (Table 4) can be classified into three groups. The first group discusses successes in applying analytics, the second describes the status of BD in PSM, and the third concerns the potential benefits that BD can have for PSM decisions. In terms of success factors in the area of PSM for BD, it is emphasized that companies need to obtain the required analyst skills (Handfield, 2014). Other factors that affect success are ensuring internal and external integration of their information systems before applying BD (Paajanen et al., 2017), including suppliers in the decision-making (Banerjee and Mishra, 2016), and knowledge about how or why the organization will use the BD results (Sanders, 2016) and apply performance measurements for BD analytics (Handfield, 2014). The studies looking at the status of analytics in the area in practice suggest that the PSM area is of growing importance (Srai and Lorentz, 2019; Sanders, 2016; Schoenherr and Speier-Pero, 2015) and that several companies are outsourcing parts of the analytical process to third parties (Sanders, 2016; Handfield, 2014). The articles on the benefits of BD either introduce an applicable model (which we will further explore in this paper), map the opinion of experts/practitioners, or investigate the link between BD and PSM outcomes by regression path analysis, and suggest benefits ranging from performance to improved PSM decisions (see Table 4 for details).

Table 4 Findings from the empirical articles (N = 20).

	Finding	**Reference**

BD application in PSM; success factors	It is essential to include suppliers in decision-making.	Banerjee and Mishra, 2016
	There is a need for market intelligence and cost modelling capabilities. SMI needs to be integrated into operational decisions.	Handfield, 2014
	Management support significantly influences adoption intention. Among the studied empirical factors, SC connectivity moderates the adoption intention.	Lai et al., 2018
	BD is critical in creating systematic SMI. Value and competitive advantage originate in the intersection of different sources of data.	Paajanen et al., 2017
	Investments in developing capabilities, resources and roles are needed.	Richey et al., 2016
	Shows how analytics can be implemented, and following maturity stages ensures data quality, refinement of key metrics, and timely access to information, and that outsourcing the technical part is a valid.	Sanders, 2016
Status of BD in practice in PSM	They show companies and applications with envisioned future applications of BD in different PSM areas including supplier identification, selection, increased spend and transaction visibility, and control and coordination.	Srai and Lorentz, 2019
	BD is most useful in providing decision support for purchasing and information for supplier negotiations. Out of the studied processes, logistics, service and planning are found to benefit most from BD, whereas sourcing, manufacturing and return management are found to benefit less.	Brinch et al., 2018
	Several companies are outsourcing the analytical part of the process. There is a need for performance measures.	Handfield, 2014
	Sourcing is the fastest growing area in the SC in terms of BD. The need for outsourcing analytics will grow as IT capabilities evolve requiring ever more specialized skills.	Sanders, 2016
	41% of the survey respondents reported having a BD application initiative and 8 % had plans to incorporate such in the future	Schoenherr and Speier-Pero, 2015
Potential benefits of BD in the area of PSM	Strategic processes related to sourcing and planning can benefit from BD	Brinch et al., 2016
	Monitoring supply can reduce supply delivery issues	Chircu et al., 2014
	Supplier selection, evaluation and price negotiation can benefit from BD	Deswal and Garg, 2015
	Collaboration, visibility and risk management can benefit from BD	Kache and Seuring, 2017
	Supplier evaluation and selection can benefit from BD	Matta and Tayal, 2017 Pouraghabagher and Sarfaraz, 2018
	Performance of internal PSM in terms of cost, time, quality and flexibility can benefit from BD	Moretto et al., 2017
	Risk and security governance, transparency can benefit from BD	Richey et al., 2016
	BD applications enable reducing uncertainty in SC as it improves forecasts, and by increasing information processing capacity it facilitates fast decision-making and responses to disruptions.	Roßmann et al., 2018
	Social media can facilitate online communication and knowledge sharing in the digital market place and provide/obtain useful feedback, which can be valuable in new product development or in risk monitoring in a supply chain network.	Swain and Cao, 2017
	Vendor rating has the potential to benefit largely from BD	Raman et al. 2018
	Supply visibility and coordination can benefit from BD	Tan and Lee, 2015
	Spend management can benefit from BD	Choi et al. 2018

2.5. Scope of PSM-related BD

The majority of the articles (i.e., 37) use the term *big data*, followed by other terms, such as, *supplier intelligence* (Jeeva and Dickie, 2012), *supply market intelligence* (Handfield, 2014), and *word of mouth data* (Swain and Cao, 2017). Others refer to the concept using phrases such as the "*huge amount of information regarding SC members*" (Martín-Rubio et al., 2016) or the "*vast amount of data in various forms*" (Sahay and Ranjan, 2008), without using the actual term *big data*. While these articles used different characterizations of BD, the majority referred to *the 3Vs*, describing *volume* (beyond traditional data management capabilities), *variety* (data in different formats and from different sources), and *velocity* (data created at a higher speed than ever; see, e.g., Rozados and Tjahjono, 2014). Some of the articles attempted to discuss BD in the PSM context.

Table 5 A summary of BD sources and some example formats of BD applied in PSM in the articles.

BD sources		**Example formats extracted**	**References**
External	Social media *(e.g. Facebook, Twitter, discussion forums)* [1, 6, 7, 9, 14, 15, 17, 23, 24, 25]	Trends, the deviant performance of suppliers, public and private perceptions, scandals, risk and disruption signals	1. Akter and Wamba, 2016; 2. Biswas and Sen, 2016; 3. Chircu et al., 2014; 4. Connaughton and Sawchuk 2014; 5. Deswal and Garg, 2015; 6. Dhurandhar et al., 2015; 7. Fan et al., 2015; 8. Handfield, 2014; 9. He et al. 2014; 10. Ivanov, 2017; 11. Jeeva and Dickie, 2012 12. Kaur and Singh, 2017; 13. Lamba and Singh, 2016; 14. Liang et al., 2016; 15. Moretto et al., 2017; 16. Mori et al., 2012; 17. Paajanen et al., 2017; 18. Richey et al., 2016; 19. Rozados and Tjahjono, 2014 20. Russo et al., 2015; 21. Sahay et al., 2008; 22. Surasvadi et al., 2017 23. Sanders, 2016; 24. Souza 2014 25. Tan et al., 2016;
	Market *(e.g. company profiles, market research companies, the stock market, electronic marketplaces, exhibitions)* [4, 8, 9, 11, 14, 15, 17, 21, 20, 22]	Tax data, competitors, supplier possibilities, economic and market trends, innovations, indicators	
	Media *(e.g. news, websites, records, reports)* [1, 7, 9, 15, 17]	Financial news on suppliers or competitors, trends	
	Public official records *(e.g. political or economic policies, reports and forecasts, risk indicators)* [2, 4, 6, 7, 11, 22, 9]	Risk indicators, extreme weather or hazards affecting sites, no-go zones, trade deals, political limitations	
	Third-party sources *(e.g. specialized firms, trade journals, consultancies)*[8, 21]	Expert input	
	Advanced search engines [6, 17]	Topic-specific data	
Internal	Internal data systems (*e.g. EDI, ERP)* [17, 18, 21]	Tables and spread sheets, operational data, historical data, spend data	
	Request for information records *(RFQs, RFIs, e-purchasing and supply management systems)* [6, 8]	Supplier-related information, supply market insight, spend data	
	Contracts [2, 4, 6, 14, 17, 19]	Pricing, terms, agreements	
	Purchasing orders [2, 7, 10,12, 13, 19]	Product, date, time, ID, quantity	
	Delivery records [2, 7, 10, 12, 13, 19]	Time, quantity, packaging	
	Supplier records *(e.g. supplier profiles, supplier records, sub-tier supplier records)* [4, 5, 6, 9, 12, 21, 25]	Performance and qualifications	
	Production records *(e.g. management systems)* [2, 10, 16, 18, 21]	Production, inventory, return	
	SC records *(e.g. SC partner records, GPS, RFIDs)* [3, 23, 25]	SC partners, flows	
	Linguistics *(e.g. opinions of decision makers)*[5]	Expert input	

In terms of variety, a multitude of different data sources and formats were suggested in the articles, reflecting the potential of BD, which can be classified based on the source of data being internal or external (see Table 5). From the external perspective, data from different social media sources, data from or on the supply market (e.g., electronic market places, market indices, market journals), data gathered from different media forms, and risk-related data (such as, indices, reports, critical commodity or supplier lists, weather reports, natural disasters, etc.) were stressed most. Among the internal data types, spend data, purchasing data (e.g., product, ID, date, time), supplier data (e.g., qualification or performance profiles), and transactional data were highlighted most. BD can be in many different formats. In terms of the variety of formats, studies discuss how storage and analysis of structured data have been the norm in most firms until recently, using languages such as SQL to

manage large data sets. However, today, more advanced analytics, on the one hand, and availability of data, on the other, are making "untagged" or unstructured information (e.g., video, free-form text, or images) accessible to decision makers.

2.6. Application of BD in PSM

A holistic process (Akter and Wamba, 2016), a collection of technologies (Chehbi-Gamoura and Derrouiche, 2017), data-driven approaches (Choi et al., 2018), data-based tools (Richey et al., 2016), or qualitative and quantitative analytics (Wang et al., 2016) are used to transform these data into insights that are not attainable through traditional analytics. The transformation of data and information into meaningful insights and support for decision-making has been highlighted in the context of BD (see, e.g., Akter and Wamba, 2016; Sanders, 2016; Nguyen et al., 2017; Choi, et al., 2018). A notable number of the articles suggested the use of an online analytical processing (OLAP) system that includes drilling, pivoting, dicing, and aggregating data (see, e.g., Biswas and Sen, 2016; Tan and Lee, 2015; Lee et al., 2009; Shaoling and Yan, 2008).

The PSM areas (i.e., expanded from the classification in Van Weele's 2010 purchasing process) that can benefit from BD are summarized in Table 6. From the total sample of our reviewed articles, PSM decisions related to supply risk management and supplier selection were mentioned most. In general, PSM areas with an operational orientation received less attention.

Table 6 Suggested PSM areas that can be supported by BD.

PSM areas	**References**
Alignment of purchasing strategies with corporate strategies [3, 5, 10, 18, 28, 39, 37, 43]	1. Akter and Wamba, 2016; 2. Ashayeri and Selen, 2008; 3. Banerjee and Mishra, 2016; 4. Barbosa et al., 2018; 5. Brinch et al., 2017; 6. Brinch, et al., 2018 7. Biswas and Sen, 2016; 8. Chircu et al., 2014; 9. Choi et al., 2018 10. Connaughton and Sawchuk, 2014; 11. Deswal and Garg, 2015; 12. Dhurandhar et al., 2015; 13. Fan et al., 2015; 14. Giannakis and Louis, 2016; 15. Handfield, 2014; 16. He et al., 2014; 17. Ivanov, 2017; 18. Jeeva and Dickie, 2012 19. Kaur and Singh, 2017; 20. Kache and Seuring, 2017; 21. Lamba and Singh, 2016; 22. Lamba and Singh 2017; 23. Lee et al., 2009 24. Liang et al., 2016; 25. Martín-Rubio et al., 2016; 26. Matta and Tayal, 2017; 27. Meriton and Graham, 2017 28. Moretto et al., 2017; 29. Mori et al., 2012; 30. Nguyen et al., 2017; 31. Paajanen et al., 2017; 32. Richey et al., 2016; 33. Rozados and Tjahjono, 2014 34. Russo et al., 2015; 35. Sahay et al., 2008; 36. Schoenherr and Speier-Pero, 2015; 37. Shaoling and Yan, 2008; 38. Surasvadi et al., 2017 39. Sanders, 2016; 40. Souza, 2014 41. Tan et al., 2016 42. Vaidyanathan and Sabbaghi, 2010 43. Wang et al., 2016 44. Yu et al., 2017 45. Zage et al., 2013; 46. Zrenka, 2019;
Market knowledge and supplier identification [1, 2, 5, 6, 7, 10, 14, 15, 17, 20, 22, 25, 28, 31, 33, 39]	
Segmentation of supplies and suppliers [39]	
Supplier selection [7, 5, 11, 19, 21, 22, 23, 25, 26, 28, 29, 30, 31, 33, 37, 39, 42, 45, 46]	
Negotiation leverage [5, 6, 7, 23, 28, 32, 36, 37, 39]	
Contracting [28, 36]	
Supplier relationship management including better coordination and integrations [1, 4, 10, 14, 25, 27, 29, 31, 32, 35, 39, 40, 44]	
Forecasting of need [10, 14, 22, 28, 32]	
Follow up and evaluation [6, 8, 23, 24, 26, 28, 31, 37, 42, 43]	
Ordering [19, 21, 32]	
Spend management [9, 10, 15, 28, 30, 38, 42]	
Monitoring and expedition [8]	
Supply risk management [10, 12, 13, 15, 16, 17, 20, 22, 24, 25, 27, 28, 30, 31, 32, 35, 39, 40, 36, 43, 45]	

Additionally, from the articles that suggest an actual application solution for BD in PSM, we identified a number of common themes (see Table 7). The three major ones are models for using BD for better supplier selection, better supply risk management (e.g., detection of fraud, abnormalities, anomalies, unusual behavior) and improved visibility for PSM (i.e., internal demand aggregation, visualization of PSM, and spend pattern identification). Chircu et al., (2014) show a case of BD improving the monitoring of flows from suppliers and thus decreasing instances of counterfeiting. The majority of the models are within the industrial context and from a large manufacturer's perspective, and three are developed within the public procurement context (on risk management and increased visibility of spend).

Table 7 The findings of the articles that suggest an application solution for BD in PSM (N = 19).

Theme	Application solution	Context	Method	Reference
Supplier selection	Various ranking procedures are applied to handle BD in supplier selection	Large manufacturing firms	Model based on survey of 235 suppliers	Matta and Tayal, 2017
	A model using artificial intelligence for a supplier selection case	Large manufacturing firm	Model based on 1 manufacturer and 18 suppliers	Deswal and Garg, 2015
	A model suggesting a selection algorithm based on how similar competitors have rated the supplier	Large manufacturing firms	Netflix open sourced data - 685,000 rows of data = ±4,700users	Pouraghabagher and Sarfaraz, 2018
	A model considering multiple products and periods, and carbon emissions, using largely varied, dynamic, and long generated data	Large manufacturing firms	Modelling: no data	Lamba and Singh, 2019
	An artificial intelligence (AI) model to select plausible partners is suggested and tested	Large manufacturing firms	Validated by experiments with 30 660 man.	Mori et al., 2012
	An analytical model to use BD for better and sustainable supplier selection and order allocation	Large manufacturing firms	Modelling: no data	Kaur and Singh, 2017
	A framework for the application of advanced analytics and OLAP in order to better select, evaluate and negotiate with suppliers	Large manufacturing firms	Modelling: no data	Lee et al., 2009
	An e-procurement system based on agent technology and OLAP in order to better select, evaluate and negotiate with suppliers	Large manufacturing firms	Modelling: no data	Shaoling and Yan, 2008
	A simulation model using BD on capacity distributions to make dual versus single sourcing decisions	Large manufacturing firms	Modelling: no data	Ivanov, 2017
Supply risk management	A tool to identify PSM-related fraud/risk using public and private sources	Public procurement	Model based on 100+ fraud cases and used on 65 000+ vendors	Dhurandhar et al., 2015
	A conceptual framework and simulation for using BD for supply risk management	Large manufacturing firms	Modelling: no data	He et al., 2014
	An automated screening method to identify instances of deception	Cross-sectoral	Modelling: no data	Zage et al., 2013
	A PSM fraud detection system that uses data crawling in social networks, company profiles and the internet to combine info on the intimacy between buyers and bidders, abnormal interactions, commodity prices and supplier credits	Large firms	Modelling: no data	Liang et al., 2016
PSM visibility and transparency	An analytical model using OLAP to use BD to improve decision-making regarding internal demand aggregation and coordination	Large manufacturing firms	Model based on the transaction data of one company between 2011-14	Tan and Lee, 2015
	The application of data visualization for the quick identification of issues and quick access, e.g. to supplier, model, part info	Large firms	Conceptual	Tan et al., 2016
	An analytical system that uses open government budget requests and PSM records to identify spend patterns	Public procurement	Modelling: no data	Surasvadi et al., 2017
	Using a fuzzy cognitive map, the paper suggests a model for prioritizing decisions regarding IT purchasing	Public procurement	Open sources public data	Choi et al., 2018

	A conceptual framework suggesting the separate analytical handling of structured and unstructured data, and the use of both real-time and storage-based intelligence (including OLAP) to create real-time operations and decisions	Large manufacturing firms	Conceptual	Biswas and Sen, 2016
Monitoring and expedition	A cloud-based system that tracks supply from suppliers using RFID tags on sensitive supply	Large firm (pharmaceutical)	Case study	Chircu et al., 2014

Among these suggested models and frameworks, eight articles were based on empirics. Tan and Lee (2015) introduce an analytical model for using BD to improve decision-making and the visibility of internal demand by aggregation and better coordination using the OLAP concept. Choi et al., (2018) use open access public sources to show how BD can help companies use internal and external information on previous procurements (by the company and competitors), and market information on the technologies, among other data, to better prioritize the type of IT system they procure. Dhurandhar et al., (2015) suggest a model for identifying PSM-related fraud/risk using public and private sources of data. The private data sources include vendor master files, RFXs, risky vendor lists, vendor bank account numbers, vendor–employee bank account matches, risky commodity lists, global clip levels, company risk reports, financial indices, and social networking data, while the public data sources are, for example, government-listed forbidden parties, country perception indices, tax haven locations, and advanced search engine searches. Mori et al., (2012), Pouraghabagher and Sarfaraz, (2018), Deswal and Garg (2015), and Matta and Tayal (2017) show the application of analytical models to handle BD in order to select suppliers. All four articles either base their models on or validate their models by empirical data. Finally, Chircu et al., (2014) show the case of tracking and tracing supply from suppliers in order to improve expedition.

3. Discussions and research agenda

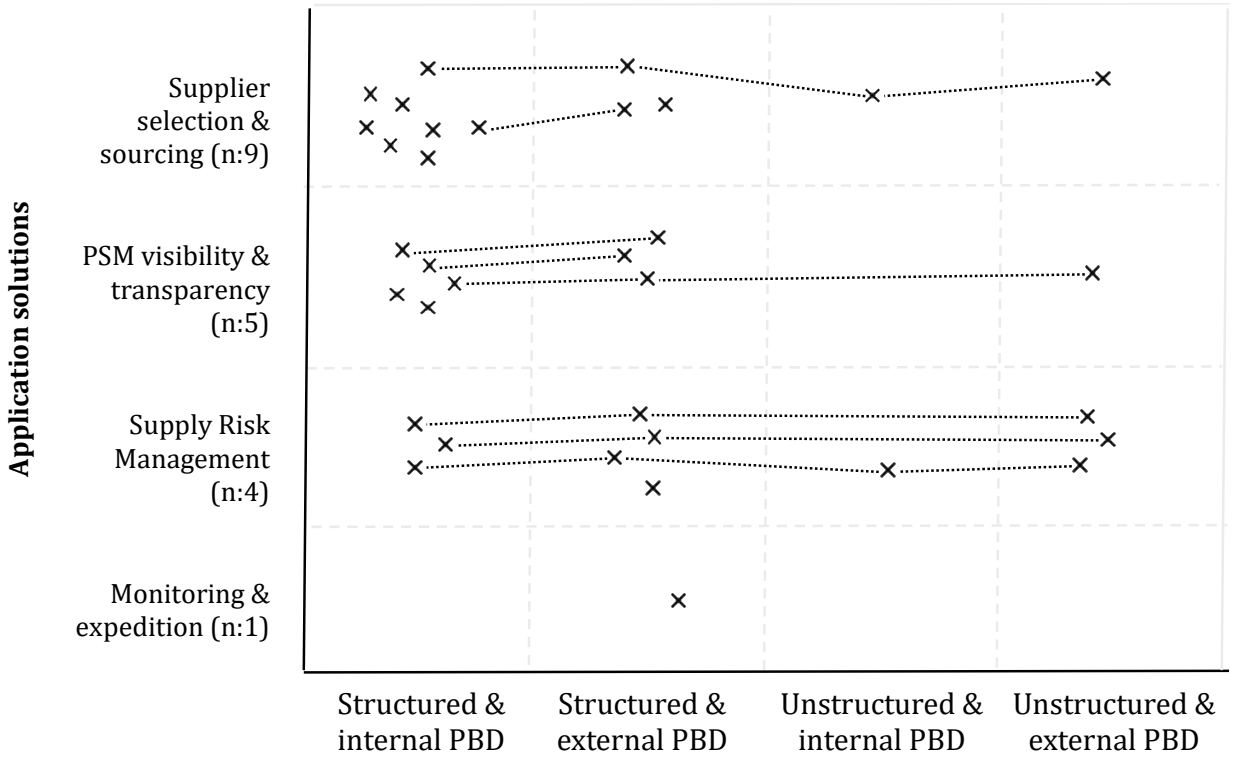
The purpose of this study was to advance knowledge generation through BD in the area of PSM through a systematic literature review. The study analyzed data sources, types of analytics and definitions; identified PSM decisions that can potentially benefit from BDA; and studies the application solutions from empirical papers. The findings summarize current understanding of how BDA can be utilized for PSM practice.

In the following sections, we will discuss the observations from the literature review to identify research needs and opportunities. The discussion includes three separate focuses: 1) use of BD in PSM related applications and introduction of a classification for PSM-BD; 2) a process application perspective; and, 3) a capabilities perspective.

3.1. Use of BD in PSM applications

Even though we identified many different data sources and formats of potential relevance to PSM, no conceptualization or categorization was found for big data sources and big data types from a PSM process perspective. The data sources and types have typically not been related to specific PSM applications. Previous research shows that regardless of the analytical tool, structured and unstructured data need to be handled separately in order to create operational decision support, as suggested by Biswas and Sen (2016). They also suggest parallel real-time and non-real-time analytical systems to handle the velocity of BD. Based on this distinction in terms of handling the

variety of formats of BD (i.e., structured versus unstructured) and the variety of sources of PSM-related BD impacting the ease of access (i.e., internal versus external), we propose PSM-related BD be classified into four groups of structured-internal (e.g., from contracts or internal data systems), structured-external (e.g., from the financial market or weather forecasts), unstructured-internal (e.g., emails or chats), and unstructured-external (e.g., social media). We mapped the application solutions in the literature based on these four PSM-BD types in order to identify research gaps and opportunities (see Figure 3).



- "n" represents the number of references suggesting an application in the given area and based on the given data
- Each cross represents the use of the data type in an application
- Dashed lines between crosses indicate the same application solution in the same paper; see Table 7 for further details

Figure 3 Mapping the suggested PSM-related application models according to the variety of source and format of PSM-BD.

Structured data are less complex to handle and access than unstructured data, which is also visible in the type of data used in the suggested application solutions in the literature. As depicted in Figure 3, the suggested application solutions in the literature mainly used structured PSM-BD, both internal (in 16 of the 17 models) and external (in 11 of the 17 models). Unstructured and external data were used in the majority of the supply risk management models (3 of the 4 models), in one model for increasing PSM visibility (out of a total of 5), and in one model for supplier selection (out of a total of 9). Unstructured and internal data were used least, and only in 2 models, and combined with other data types (one in supplier selection using linguistics to explicate internal decision maker preferences of suppliers; and one in supply risk management using internal communication channels). Subsequently, we can conclude that structured internal data are the least

complex of the four to use in PSM decisions, in terms of access and analysis, and unstructured-internal data the most complex. Unstructured internal data may be the most complex, as they appear to be the least discussed in the literature in terms of what they can entail.

As is visible in our data, of the total sample of 65 articles, only 17 suggested an application solution, of which only 7 used empirics, and of which only 2 used multiple PSM-BD types. Of the 17 models, 10 used only one type of data, of which 7 used structured and internal data (6 for supplier selection and 1 for visibility) and 3 structured and external data (for supplier selection, risk management, and expedition). Of the models, 3 used both internal and external structured data, 3 combined unstructured external data with structured formats, and 2 used a combination of all four types of PSM-BD (i.e., in supplier selection and supply risk management).

These observations suggest that the analytics of simultaneous handling of structured and unstructured data are much more complicated than combining structured formats from varied sources. This can be explained by the fact that the two formats need to be handled separately in, e.g., parallel systems (e.g., as also contended by Biswas and Sen, 2016). The use of empirics in several of the application solutions shows the practical relevance of the topic, and the application models are mainly developed within the industrial context of large manufacturers, with a few also focusing on the public procurement context (specifically on spend management and risk management). However, in relation to specific PSM application areas, current literature does not provide answers regarding which data source can support specific PSM decision areas and how it can do so. Furthermore, it is unclear if the connection between these concepts, or the nature of each of them, would change in different contexts (e.g., it is unclear if a certain social media-based data source is more or less suitable for a specific industry or PSM decision). This calls for more research to guide practitioners and researchers in specific PSM areas from two angles (which we will elaborate on in the coming section of this paper):

1. The sources, types, and formats of PSM-BD for specific PSM application areas in specific contexts

For instance, we need more clarity on what type of PSM-BD (within the structured/unstructured and internal/external groups) would be more or less suitable for decisions regarding the PSM application areas with high expected potentials but less research on the type of data, including those identified in this literature review (see Table 7), supplier evaluation and selection, supply market knowledge, spend, contract and category management, and monitoring and expedition.

2. The analytics of using unstructured and combination PSM-BD for application areas with high expected potentials

Such areas include those with fewer developed application models on how to combine different PSM-BD types but with potential interest, e.g., those identified in our literature review (see Tables 4 and 7) on advanced spend management, supplier evaluation and selection, supply risk management, supply monitoring, and supply market knowledge. Next, we will elaborate on such research needs in order to advance BD in the PSM practice.

3.2. Process application perspectives

The literature identifies possible PSM application areas for BD, with the majority of studies being related to supplier selection/evaluation and supply risk management. We believe these will be important areas in future studies. For *supplier selection*, prescriptive models for handling PSM-BD are suggested (i.e., Deswal and Garg, 2015; Matta and Tayal, 2017). The majority only use internal

and structured PSM-BD, and only two of these studies use empirical data. The developed models often simplify supplier selection to merely include the choice of a suitable partner from market potentials. Supplier selection, however, spans from market scanning and analysis to identifying potential partners, to evaluating long lists of suppliers to create shorter lists and deeper evaluation of short-listed suppliers as an input to negotiation or contracting (van Weele, 2010). Thus, we suggest more research around the following two questions:

- *What are the different types of structured and unstructured data from internal and external sources that can help better identify potential suppliers and evaluate suppliers as input for negotiations and supplier selection?*
- *How can suitable structured and unstructured data from internal and external company sources be combined to assist identification of potential suppliers and evaluation of suppliers as input for negotiations and supplier selection?*

Supply risk management is a relatively advanced area in the studied literature, with at least three models showing the possibility of prescriptive and predictive models using both structured and unstructured data from internal and external sources to better detect and, thus, manage disruptions (Zage et al., 2013; Dhurandhar et al., 2015; Liang et al., 2016). However, external supplier data from the supplier companies can be of differing accuracy (e.g. between risky and top performing suppliers) making the issue of big data verification essential in this area. Additionally, the area of supply risk management includes a variety of issues connected to risk sources and possible solutions, which are not addressed in the existing studies. Future research should build on the existing PSM-BD models and supply risk management studies and answer the question:

- *How can advanced analytics verify the accuracy of various big data sources and format on the risky suppliers compared to the top performing suppliers for risk management?*
- *What are the topics and areas in supply risk management that BD can support or solve?*

Utilizing spend data is another area with potential for research, which mostly includes post-contract structured internal data. In our study, we identified models aimed at increasing spend in order to develop support for, e.g., supply and supplier choice. These data are easily accessible and include relevant information about activities close to the company. The spend management decision is also closely connected to *category management*, which is, in turn, is tightly connected to supply market analysis (van Weele, 2010). More studies are needed to understand:

- *How structured spend data can be integrated with external data sources to assist market analysis and category management?*

Spend data usage can also be widened beyond the single company perspective. All the reviewed literature on BD in PSM in the private sector take the viewpoint of a single company and concentrate on data collected by and in relation to a specific company and its PSM activities. We suggest cross-company or cross-industry studies on multi-company data, focusing on a certain PSM process or topic. Such cross-organization spend management analyses are more common in public procurement, as the purpose is to look over multiple public entities with organizational borders (e.g., Choi et al., 2018). In the private sector, a third-party organization can provide benchmarking data with a similar approach. Previous research, such as Bakker et al., (2008), discuss both virtual and third-party structures useful for governing interorganizational relations and shared information platforms. Related analysis could target, for example, indirect spend, expected to be comparable across industries and companies. Contracts offer another database that includes much structured data and can offer a valuable source of information. Consequently, one question to be addressed is:

- *How could structured data from databases that include multiple companies provide insight for PSM decision areas?*

A PSM area that is not gaining much attention in the current BD literature, but one that is obvious from the viewpoint of BD, is *supply market intelligence* (SMI), in other words, the capability to develop valuable knowledge about supply markets specifically for the purpose of supporting accurate and confident decision-making in the procurement process (Handfield, 2014). Specifically, here, where the focus is mostly on pre-contract topics, the potential benefits of BD and analytics are big. Along with the increasing amount of data and information in the PSM function, the need for efficient SMI processes and routines and their integration into PSM decision-making becomes increasingly obvious. Thus, further research could provide answers to the question:

- *How could SMI processes and capabilities be developed to best support PSM decision-making?*

3.3.Capabilities perspective

The current literature typically takes the viewpoint of the potential benefits of BD in various business settings (see, e.g., Tiwari et al., 2018). Very few studies have discussed firm-level capability and maturity in their conceptualizations or models. These two aspects are important in the application of BD for business and should be studied independently in further studies. As previous research has noted, "*an intention must be coupled with practical knowledge in order to transform into action',* and firm capability is found to be pivotal to the adoption of system and technological solutions (Brock and Khan, 2017). In fact, many researchers identified learning capabilities as the most important component for innovation and adoption of new processes as well as technologies.

The required capabilities for the application of BD in PSM also need greater understanding. There are some studies on BD capabilities (Akter and Wamba, 2016; Arunachalam et al., 2018), but specific studies on PSM are non-existent. The strategic role of PSM, the impact of the PSM decisions on other functions, and the performance of the whole company make capability development essential in order to strengthen the role of BD in this area. Studies on this subject should explore how the management, technology, and talent related to BD capabilities contribute to PSM. These studies could build on the findings of BD capability studies in other domains. Consequently, we suggest future research on:

- *What BD capabilities are needed in the area of PSM and how can they be organized to improve functional and firm-level performance?*

More data sources and unstructured data formats could potentially be relevant to use in pre-contract applications, when it is not yet obvious which objects (suppliers, items, etc.) to focus on, such as in SMI (e.g. Selen and Ashayeri, 2008). In post-contract applications, however, there are more available data on existing suppliers, items, and deliveries. There is also a difference between the analytics possibilities regarding pre-contract and post-contract decisions. Pre-contract decisions are less defined and the questions to be asked are broad. Post-contract decisions are more defined, the efforts needed are more focused, and they can be accessed from company databases or the supply chain. BD can be expected to differ between pre- and post-contract applications, which motivates the following question for future research:

- *Do BD capabilities differ between pre- and post-contract applications and, if so, how?*

4. Concluding remarks, contributions and limitations

This literature review provides a state-of-the-art description of the documented knowledge of BD in PSM. The analysis shows that few published studies on BD take the perspective of PSM. In accordance with other literature reviews on BD in the supply chain management area (e.g. Wang et al., 2016; Nguyen et al., 2017), ours shows that the number of publications in the PSM area is increasing, indicating a growing interest in the topic academically. Thus, our findings contribute to the PSM area by *firstly*, depicting the current knowledge about the research in, and the application areas of, BD in PSM. With the systematic literature review, we identify all the possible studies conducted on the topic, which form a frame of reference for future researchers and managers to identify relevant studies. The study analyses BD sources, types of analytics; definitions; identifies PSM decisions that can potentially benefit from BD; and studies application solutions from empirical papers. The findings summarize current understanding of how BD can be utilized for PSM practice. In general, the area seems to be immature, as literature provides limited understanding of the matter, relies largely on conceptual studies, and few empirical studies aiming at theory building were identified. This literature review provides a state-of-the-art description of the documented knowledge of BD in PSM. The analysis shows that it is a premature research area and few published studies on BD take the perspective of PSM. In accordance with other reviews on BD, our review shows that the number of publications in the area is increasing, indicating a growing interest in the topic academically. Additionally, the survey-based studies (e.g. Schoenherr and Speier-Pero, 2015; Sanders, 2016; Srai and Lorentz, 2019) show that this area is growing within practice.

Secondly, we contribute to the BD discussions in the PSM area by better defining what PSM-BD is. We introduce a classification for this data and bring several examples of such data introduced in the literature. Accordingly, PSM-BD can fall within either of the four groups of structured-internal (e.g. from internal data systems of contracts), unstructured-internal (e.g. emails and chats), structured-external (e.g. market indices and public office records), and unstructured-external (e.g. media, expert opinions, social media).

Thirdly, our findings contribute to research on BD in PSM by identifying different research needs and directions. Based on the synthesis of the extant literature we develop a set of research questions in need of further research to guide future studies. The area in general needs conceptualization and theory building in the form of empirical studies. The majority of the BD applications in the PSM area are currently conducted using internal and structured data, but the expectations of scholars for the near future are much more application usage focused on using more complex mixes of structured and unstructured internal and external data in strategic PSM decisions, such as supplier selection and relationship management, or supply risk management (Lamba and Singh, 2017; Nguyen et al., 2017). In our study, we found such application solutions in three PSM areas of supplier selection (Mori et al., 2012; Deswal and Garg, 2015; Pouraghabagher and Sarfaraz, 2018), supply risk management (Dhurandhar et al., 2015; Liang et al., 2016), and improved PSM visibility on, e.g., spend data (Choi et al., 2018).

Consequently, we *fourthly*, contribute to these three specific PSM areas of supply risk management, supplier selection, and PSM visibility, by identifying how future research on BD can improve decision making and BD related application solutions in these areas. Supply risk management was the most mature area that we identified, with developed application models using various BD

sources and formats. We argue that this maturity is related to the higher clarity of the four types of BD for that purpose. In other application areas, extant research does not provide answers regarding which BD source can support each PSM decision and how it can do so. This calls for more research to guide practitioners and researchers in specific PSM areas: a) on the sources, types, and formats of BD for specific PSM application areas, and, b) on the analytical side of using unstructured data and combination data for application areas with high expected potentials, including advanced spend management, supplier evaluation and selection, supply risk management, category management, and supply market knowledge. *Fifthly*, we also propose that the pre-contract phase of a PSM process, when the questions to be answered are less focused and more open-ended, is best supported by the use of variable and many data sources, and predictive and prescriptive analytics, which needs development of analytical solutions to handle the variety and dynamics and data formats.

Some managerial implications can be extracted from the findings. First, PSM is a growing area in which to apply BD, and efforts towards developing it are justified. One of the challenges in this area is connected to adopting the analytics technology needed, as it may not be the core capability of a company. Therefore, outsourcing this part of the analytics is a valid option. However, the company needs to develop assimilation capabilities in-house–and in our case, in particular in PSM – in order to ensure the efficient integration of the provided intelligence into decision-making. Finally, it is important to note that this study and its findings are limited to the identified academic articles and based on the keywords used. This review only covers the findings based on this sample in this fast-developing area. Our observations are also limited to our coding method and focused on the application use of BD in PSM, and thus topics such as challenges in the utilization or acquisition of big data in PSM are not extracted from the literature.

References

Akter, S. and Wamba, S.F. (2016), 'Big data analytics in e-commerce: A systematic review and agenda for future research', *Electronic Markets*, Vol. 26 No. 2, pp. 173–194.

Arunachalam, D., Kumar, N., and Kawalek, J.P. (2018), 'Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, pp. 416–436.

Bakker, E., Walker, H., Schotanus, F., & Harland, C. (2008). 'Choosing an organisational form: the case of collaborative procurement initiatives', *International Journal of Procurement Management,* Vol. 1, No. 3, pp. 297-317.

Banerjee, M. and Mishra, M. (2017), 'Retail supply chain management practices in India: A business intelligence perspective', *Journal of Retailing and Consumer Services*, Vol. 34, pp. 248–259.

Barbosa, M.W., de la Calle Vicente, A., Ladeira, M.B., and de Oliveira, M.P.V. (2018), 'Managing supply chain resources with big data analytics: A systematic review', *International Journal of Logistics Research and Applications*, Vol. 21 No. 3, pp. 177–200.

Biswas, S. and Sen, J. (2016), A proposed architecture for big data driven supply chain analytics', *ICFAI University Press (IUP) Journal of Supply Chain Management,* Vol. 13 No. 3, pp. 7–34.

Brinch, M., Stentoft, J., and Jensen, J.K. (2017), 'Big data and its applications in supply chain management: Findings from a Delphi study', In *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1351–1360.

Brinch, M., Stentoft, J., Jensen, J.K., and Rajkumar, C. (2018), 'Practitioners understanding of big data and its applications in supply chain management', *The International Journal of Logistics Management*, Vol. 29 No. 2, pp. 555–574.

Chehbi-Gamoura, S. and Derrouiche, R. (2017), 'Big valuable data in supply chain: Deep analysis of current trends and coming potential', In Camarinha-Matos, L., Afsarmanesh, and H., Fornasiero, R. (eds.) *Collaboration in a Data-Rich World. PRO-VE 2017. IFIP Advances in Information and Communication Technology*, Vol. 506, Springer, Cham, pp. 230–241.

Chircu, A.M., Sultanow, E., and Chircu, F.E. (2014), 'Cloud computing for big data entrepreneurship in the supply chain: Using SAP HANA for pharmaceutical track-and-trace analytics', In *2014 IEEE 10th World Congress on Services*, Jun. 27-Jul. 2, 2014, Anchorage, AK, USA, pp. 450–451.

Choi, Y., Lee, H., and Irani, Z. (2018), 'Big data-driven fuzzy cognitive map for prioritising IT service procurement in the public sector', *Annals of Operations Research*, pp. 1–30.

Connaughton, P. and Sawchuk. C. (2014), 'Procurement analytics in the era of big data: What CPOs need to know', The Hackett Group. http://www.planetic.es/sites/default/planeticfiles/contentfiles/private/The%20Hackett%20Group%20Analytics%20in%20the%20Era%20of%20Big%20Data%20(Octubre2014).pdf.

Deswal, M. and Garg, S.K. (2015), 'Application of topsis: A multiple criteria decision making approach in supplier selection', In *2nd International Conference on Recent Innovation in Science, Engineering and Management*, Jawaharlal Nehru University, New Delhi, India.

Dhurandhar, A., Graves, B., Ravi, R., Maniachari, G., and Ettl, M. (2015), 'Big data system for analyzing risky purchasing and supply management entities', In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1741–1750.

Fan, Y., Heilig, L., and Voß, S. (2015), 'Supply chain risk management in the era of big data", In Marcus, A. (ed.), *Proceedings (Part I) of the 4th International Conference, DUXU 2015*, Los Angeles, CA, USA, Aug. 2-7, pp. 283–294.

Giannakis, M. and Louis, M. (2016), 'A multi-agent based system with big data processing for enhanced supply chain agility', *Journal of Enterprise Information Management*, Vol. 29 No. 5, pp. 706–727.

Handfield, R. (2014), 'Organizational structure and application of supply market intelligence: Current trends and best practice', In *Proceedings of the 2014 Workshop on Human Centered Big Data Research*, pp. 36–40.

He, M., Ji, H., Wang, Q., Ren, C., and Lougee, R. (2014), 'Big data fueled process management of supply risks: Sensing, prediction, evaluation and mitigation', In Tolk et al., *Proceedings of the 2014 Winter Simulation Conference*, pp. 1005–1013.

Huang, S.-F., Su, C.J., and Saballos, M.B.V. (2018), 'Social media big data analysis for global sourcing realization', In Bhatia, S., Mishra, K., Tiwari, S., and Singh, V. (eds.) *Advances in*

Computer and Computational Sciences. Advances in Intelligent Systems and Computing, Vol. 554, Springer, Singapore.

Ivanov, D. (2017), 'Simulation-based single vs. dual sourcing analysis in the supply chain with consideration of capacity disruptions, big data and demand patterns', *International Journal of Integrated Supply Management*, Vol. 11 No. 1, pp. 24–43.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J., Ramakrishnan, R., and Shahabi, C. (2014), 'Big data and its technical challenges', *Communications of the ACM*, Vol. 57 No. 7, pp. 86–94.

Jeeva, A.S. and Dickie, C. (2012), 'A taxonomic approach to supplier intelligence in manufacturing: Managing components of strategic purchasing and supply management planning', *International Journal of Business Environment*, Vol. 5 No. 1, pp. 88–100.

Kache, F. and Seuring, S. (2017), 'Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management', *International Journal of Operations & Production Management*, Vol. 37 No. 1, pp. 10–36.

Kamann, D. J. F., Alfaro, J. A., & Sebrek, S. S. (2019) '28 years of traceability management: trends, bottlenecks and opportunities', *International Journal of Procurement Management*, Vol. 12, No. 1, pp. 56-87.

Kwon, O., Lee, N., and Shin, B. (2014), 'Data quality management, data usage experience and acquisition intention of big data analytics', *International Journal of Information Management*, Vol. 34, No. 3, pp. 387–394.

Lai, Y., Sun, H., and Ren, J. (2018), 'Understanding the determinants of big data analytics (BD) adoption in logistics and supply chain management: An empirical investigation', *The International Journal of Logistics Management*, Vol. 29 No. 2, pp. 676–703.

Lamba, K. and Singh, S.P. (2018), 'Dynamic supplier selection and lot-sizing problem considering carbon emissions in a big data environment', *Technological Forecasting and Social Change*, Vol. 144, pp. 573-584.

Lamba, K., Singh, S.P., and Mishra, N. (2018). "Integrated decisions for supplier selection and lot-sizing considering different carbon emission regulations in big data environment', *Computers & Industrial Engineering*, Vol. 128, pp. 1052-1062.

Lee, C.K.M., Lau, H.C.W., Ho, G.T.S., and Ho, W. (2009), 'Design and development of agent-based procurement system to enhance business intelligence', *Expert Systems with Applications*, Vol. 36, pp. 877–884.

Liang, X., Ye, Z., Fan, Y., and Yin, K. (2016). "Procurement fraud detection based on big data and social network analysis', *Journal of Residuals Science & Technology*, Vol. 13 No. 7, pp. 146.1–146.6.

Martín-Rubio, I., Tarquis, A.M., and Andina, D. (2016), 'Business intelligence: New products development and supply chain systems in a SoSE perspective', In *World Automation Congress (WAC), 2016*, IEEE, pp. 1–5.

Matta, P. and Tayal, A. (2016), 'Supplier performance evaluation for manufacturing industries: Re-exploring with big data analysis', In *International Conference on Advances in Computing and Data Sciences*, Springer, Singapore, pp. 516–526.

Meriton, R. and Graham, G. (2017), 'International supply chain resilience: A big data perspective', In *Globalisation 2.0 Symposium Proceedings. 21st Cambridge International Manufacturing Symposium–Globalisation 2.0,* Sep. 28-29. Cambridge, UK. pp. 158–164.

Moretto, A., Ronchi, S., and Patrucco, A.S. (2017), 'Increasing the effectiveness of purchasing and supply management decisions: The value of big data in the purchasing and supply management process', *International Journal of RF Technologies,* Vol. 8 No. 3, pp. 79–103.

Mori, J., Kajikawa, Y., Kashima, H., and Sakata, I. (2012), 'Machine learning approach for finding business partners', *Expert Systems with Applications,* Vol. 39 No. 12, pp. 10402–-07.

Nguyen, T., Zhou, L., Spiegler, V.L.M., Ieromonachou, P., and Lin, Y. (2017), 'Big data analytics in supply chain management: A state-of-the-art literature review', *Computers & Operations Research,* Vol. 98, pp. 254-264.

Paajanen, S., Valkokari, K., and Aminoff, A. (2017), 'The opportunities of big data analytics in supply market intelligence', In Camarinha-Matos, L.M. et al., (eds.), *PRO-VE 2017, IFIP AICT 506,* Springer, pp. 194-205.

Research Methods, Sage Publications, Thousand Oaks, CA, pp. 671–689.

Richey Jr, R.G., Morgan, T.R., Lindsey-Hall, K., and Adams, F.G. (2016), 'A global exploration of Big Data in the supply chain', *International Journal of Physical Distribution & Logistics Management,* Vol. 46 No. 8, pp. 710–739.

Roßmann, B., Canzaniello, A., von der Gracht, H., and Hartmann, E. (2018), 'The future and social impact of big data analytics in supply chain management: Results from a Delphi study', *Technological Forecasting and Social Change,* Vol. 130, pp. 135–149.

Rousseau, D.M., Manning, J., and Denyer, D. (2008), '11 evidence in management and organizational science: Assembling the field's full weight of scientific knowledge through syntheses', *Academy of Management Annals,* Vol. 2 No. 1, pp. 475–515.

Rozados, I.V. and Tjahjono, B. (2014), 'Big data analytics in supply chain management', In *6th International Conference on Operations and Supply Chain Management,* Bali. doi: 10.13140/RG.2.1.4935.2563

Russo, I., Confente I., and Borghesi, A. (2015), 'Using big data in the supply chain context: Opportunities and challenges', In Garlatti, A. and Mazzaro, M. (eds.), *Proceedings of The 16th European Conference on Knowledge Management ECKM 2015,* Udine, Italy.

Sahay, B.S. and Ranjan, J. (2008), 'Real time business intelligence in supply chain analytics', *Information Management & Computer Security,* Vol. 16 No. 1, pp. 28–48.

Sanders, N.R. (2014), 'Big data driven supply chain management: A framework for implementing analytics and turning information into intelligence', Pearson Education Inc, Upper Saddle River, New Jersey.

Sanders, N.R. (2016), 'How to use big data to drive your supply chain', *California Management Review,* Vol. 58 No. 3, pp. 26–48.

Selen, W., & Ashayeri, J. (2008) 'Global sourcing strategy alignment using business intelligence: a conceptual framework', *International Journal of Procurement Management,* Vol. 1, No. 3, pp. 342-358.

Shaolin, D. and Yan, L. (2008), 'Design of an e-procurement system based on business intelligence tools', In *Proceedings of The International Conference on Management of e-Commerce and e-Government*, pp. 324–327.

Schoenherr, T. and Speier-Pero, C. (2015), 'Data science, predictive analytics, and big data in supply chain management', *Journal of Business Logistics*, Vol. 36 No. 1, pp. 120–132.

Srai, J. S., & Lorentz, H. (2019) 'Developing design principles for the digitalisation of purchasing and supply management', *Journal of Purchasing and Supply Management*, Vol. *25*, No. 1, pp. 78-98.

Stentoft, J., Mikkelsen, O.S. and Brinch, M., (2016), 'Big data applications in sourcing processes', *Dilf Orienteering*, Vol. 53 No. 4, pp. 38–42.

Souza, G.C. (2014), 'Supply chain analytics', *Business Horizons*, Vol. 57 No. 5, pp. 595–605.

Surasvadi, N., Saiprasert, C. and Thajchayapong, S. (2017), 'Budget and procurement analytics using open government data in Thailand', *2017 10th International Conference on Ubi-media Computing and Workshops*, Ubi-Media, Pattaya, pp. 1–6. doi: 10.1109/UMEDIA. 2017.8074079

Swain, A.K. and Cao, R.Q. (2017), 'Using sentiment analysis to improve supply chain intelligence', *Information Systems Frontiers*, pp. 1–16.

Tan, H. and Lee, W.L. (2015), 'Evaluation and improvement of purchasing and supply management process with data analytics', *International Journal of Advanced Computer Science and Applications*, Vol. 6 No. 8, pp. 70–80.

Tiwari, S., Wee, H.M., and Daryanto, Y. (2018), 'Big data analytics in supply chain management between 2010 and 2016: Insights to industries', *Computers & Industrial Engineering*, Vol. 115, pp. 319–330.

Tranfield, D., Denyer, D., and Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, Vol. 14 No. 3, pp. 207–222.

Vaidyanathan, G. and Sabbagi, A. (2010), 'Supply chain intelligence and value creation: A framework', *Issues in Information Systems*, Vol. 11 No. 1, pp. 571–576.

Van Weele, A.J. (2010), *Purchasing and Supply Chain Management: Analysis, Strategy, Planning and Practice*, Cengage Learning EMEA.

Wang, G., Gunasekaran, A., Ngai, E.W., and Papadopoulos, T. (2016), 'Big data analytics in logistics and supply chain management: Certain investigations for research and applications', *International Journal of Production Economics*, Vol. 176, pp. 98–110.

Yu, J., Chen, Y., Jia, X., and Niu, Z. (2017), 'Study on the construction of big data platform for medical supply chain', *Revista de la Facultad de Ingeniería U.C.V.*, Vol. 32 No. 5, pp. 225–229.

Zage, D., Glass, K., and Colbaugh, R. (2013), 'Improving supply chain security using big data', In *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference*, pp. 254–259.

Zrenka, W. (2017), 'The use and the future of big data analytics in supply chain management', *Research in Logistics & Production*, Vol. 7 No. 2, pp. 91–102.