EDITORIAL

DATA SCIENCE, PREDICTIVE ANALYTICS, AND BIG DATA: A REVOLUTION THAT WILL TRANSFORM SUPPLY CHAIN DESIGN AND MANAGEMENT

This article is forthcoming: Waller, M.A. and Fawcett, S.E. (2013) Data Science, Predictive Analytics, and Big Data: A Revolution that Will Transform Supply Chain Design and Management," Journal of Business Logistics, Vol. 34[2]

Matthew A. Waller
Stanley E. Fawcett

ABSTRACT

We illuminate the myriad of opportunities for research where supply chain management intersects with data science, predictive analytics, and big data, collectively referred to as DPB. We show that these terms are not only becoming popular but are also relevant to supply chain research and education. Data science requires both domain knowledge and a broad set of quantitative skills, but there is a dearth of literature on the topic and many questions. We call for research on skills that are needed by SCM data scientists and discuss how such skills and domain knowledge affect the effectiveness of a SCM data scientist. Such knowledge is crucial to developing future supply chain leaders. We propose definitions of data science and predictive analytics as applied to supply chain management. We examine possible applications of DPB in practice and provide examples of research questions from these applications, as well as examples of research questions employing DPB that stem from management theories. Finally, we propose specific steps interested researchers can take to respond to our call for research on the intersection of supply chain management and DPB.

Key Words: Data science, predictive analytics, big data, logistics, supply chain management, design, collaboration, integration, education.

INTRODUCTION

"Big data" is the buzzword of the day. But more than the typical faddish fuzz, big data carries with it the opportunity to change business model design and day-to-day decision-making that accompany emerging data analysis. This growing combination of resources, tools, and applications has deep implications in the field of supply chain management, presenting a doozy of an opportunity and a challenge to our field. Indeed, more data has been recorded in the past two years than in all of previous human history[1]. Big data is being used to transform medical practice, modernize public policy, and

[1]Source: IBM, http://www-01.ibm.com/software/data/bigdata/ accessed March 27, 2013.

inform business decision-making (Mayer-Schonberger and Cukier, 2013). Big data has the potential to revolutionize supply chain dynamics.

The growth in the quantity and diversity of data has led to data sets larger than is manageable by the conventional, hands-on management tools. To manage these new and potentially invaluable data sets, new methods of data science and new applications in the form of predictive analytics, have been developed. We will call this new confluence of data science, predictive analytics, and big data (DPB).

Data is widely considered to be a driver of better decision-making and improved profitability, and this perception has some data to back it up. Based on their large-scale study, McAfee and Brynjolfsson (2012) note, "[t]he more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results … companies in the top third of their industry in the use of data-driven decision making were on average, 5% more productive and 6% more profitable than their competitors" (p. 64). To make the most of the big-data revolution, supply chain researchers and managers need to understand and embrace DPB's role and implications for supply chain decision-making.

DATA SCIENCE, PREDICTIVE ANALYTICS AND BIG DATA

There is growing popular, business, and academic attention to DPB. For instance, the October 2012 issue of *Harvard Business Review* contained three articles that are relevant to this editorial: "Big Data: The Management Revolution" (McAfee and Brynjolfsson 2012), "Data Scientist: The Sexiest Job of the 21st Century" (Davenport and Patil 2012), and "Making Advanced Analytics Work for You" (Barton and Court 2012). MIS Quarterly had a special issue on business intelligence and the lead article was titled, "Business Intelligence and Analytics: From Big Data to Big Impact" (Chen, Chiang, and Storey 2012). There is also a plethora of articles in trade and even lay publications on these topics. There is even a new journal, *Big Data*, which premiered in March 2013.

Over the past few years, we have been trying to understand the DPB's implications for research and education in business logistics and supply chain management. We believe that these new tools will transform the way supply chain are designed and managed, presenting a new and significant challenge to logistics and SCM. Meeting this challenge may require changes in foci of research and education. Many traditional approaches will need to be re-imagined. Some standard practices may even be discarded as obsolete in the new data-rich environment. Some may see the possibilities as threats rather than opportunities. Yet DPB and SCM are fundamentally compatible, thus the tremendous value of DPB lies within our grasp.

We want to encourage submission of research on topics related to DPB that is relevant to logistics and supply chain management. Importantly, because there is a lack of agreement regarding the meanings of these terms, and because there is a dearth of articles on how these terms apply to the logistics and supply chain disciplines, we would like to facilitate the process by suggesting definitions, concepts, and avenues for research.

Data Science: Powerful Tools Made Relevant by Domain Knowledge

Generally, data science is the application of quantitative and qualitative methods to solve relevant problems and predict outcomes. One of the salient revelations of today, with the vast and growing amount of data, is that domain knowledge and analysis cannot be separated. This is another motivation to write this editorial. Research in the area of DPB is needed by researchers with domain knowledge in logistics and supply chain management. Professor Jeff Stanton of Syracuse University was quoted by Dumbill et al. (2013):

> "From a teaching perspective, as a faculty member I can teach someone how to do a t-test in 10 minutes, and I can teach them how to write a Python program in half an hour, but what I cannot teach them very easily is the domain knowledge. In other words, in a given area, if you are from healthcare, what you need to know in order to be effective at analysis is very different than if you are in retail. That underlying domain knowledge, to be able to have a student come up to speed on that is very hard." (p. 22)

Likewise, Shelly Farnham of Microsoft was quoted by Dumbill et al. (2013):

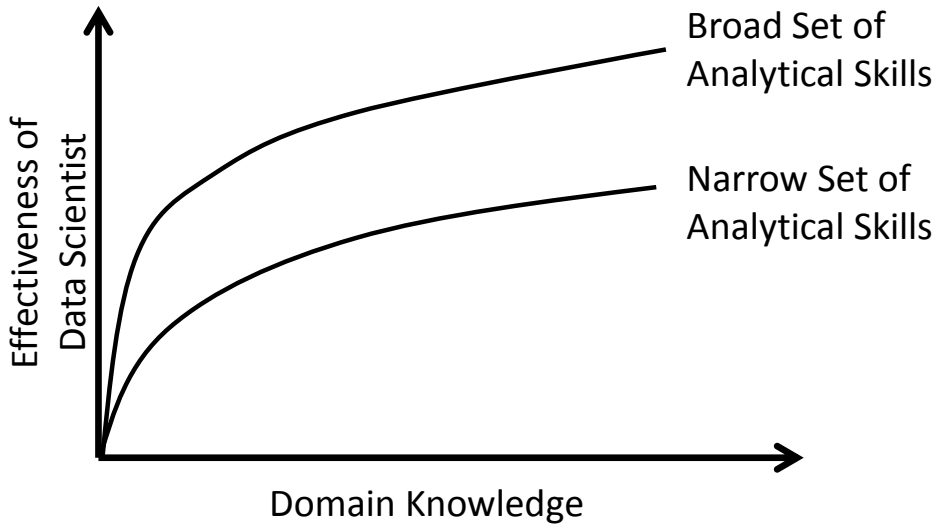
> "One of the challenges is that data science is not agnostic of domain. For example, when we are looking for people, interns or full-time people on our team, we definitely look for people who have experience analyzing data, but they also should be deeply engaged with the topic … I think that the domain knowledge is a very important aspect of what we are looking for." (p. 25)

Thus, academic and applied professionals must have both the analytical skills and the business and management understanding. As Provost and Fawcett (2013) write, "But data science involves much more than just data-mining algorithms. Successful data scientists must be able to view business problems from a data perspective." (p. 52)

Data scientists need deep domain knowledge and a broad set of analytical skills. Developing a broad set of analytical skills requires consistent investments of time. Developing deep domain knowledge requires similar dedication of effort. To that end, typically there is no single individual that can possibly have all of what is needed by a data scientist. If you have someone with lots of domain experience but limited analytical capabilities, it may be difficult to acquire the analytical capabilities. On the other hand, someone with strong analytical capabilities may not be willing to learn the domain or may never have the opportunity to learn the domain. Perhaps, developing analytical capabilities may be easier at a younger age whereas learning the intricacies of a domain may depend on accumulated experience, which emerges from motivation and well-invested time.

Although we do not have a strong body of evidence, our conjecture is that domain knowledge is necessary for data science but that the returns on domain knowledge are diminishing and that the relationship is moderated by the breadth of the analytical skills. Figure 1 illustrates the conceptual relationship between effectiveness, domain knowledge and breadth of analytical skill set. The effectiveness of a data scientist might be measured by the size of the actionable opportunities they discover.

Figure 1: Effectiveness, Domain Knowledge, and Breadth of Analytical Skill Set



However, some analytical skills are more important than others. Table 1 provides examples of skills that are needed by a data scientist in supply chain management. What is interesting is that education in most of these disciplines progresses toward more focus and less breadth. This is not only true of the quantitative disciplines, but it is also true of the functional business disciplines. Training a data scientist would require a functional business discipline to inculcate toward more breadth, rather than more depth, as a student progresses in her educational path. This would be true for the training of both the practitioner and the researcher. Perhaps there should be two paths for Master and PhD degrees in SCM, one for domain knowledge creation and dissemination and one for SCM data science[2].

Table 1: Examples of skills needed by a SCM data scientist

Discipline	SCM Data Scientist Skill Set	
	More Important	**Less Important**
Statistics	Broad *awareness* of many different methods of estimation and sampling	Derivations of methods and proofs of maximum likelihood estimation
Forecasting	Understanding *application* of qualitative and quantitative methods of forecasting	Understanding of underlying stochastic processes
Optimization	Numerical methods of optimization	Finding global optimal solutions
Discrete event	Quick design and implementation of	Queuing theory

[2] It is tempting to think that this is the difference between marketing and marketing science, however, most PhDs in marketing science tend to have a very strong focus on one methodological area.

simulation	discrete event simulation models	
Applied probability	Using probability theory with actual data to estimate the expected value of random variables of interest	The theory of stochastic processes
Analytical mathematical modeling	Using numerical methods to estimate functions relating independent variable to dependent variables	Proving theorems
Finance	Capital budgeting	Efficient market theory
Economics	Determining opportunity cost	Macroeconomic theory
Marketing	Marketing science	Semiotics
Accounting	Managerial accounting	Debits and credits journal entries

You will probably notice the applied nature of the skills needed by an SCM data scientist in Table 1. However, this does not mean that a strong theoretical education is not needed in SCM. In fact, a strong theoretical knowledge is crucial, *within the area of SCM*. That is, a SCM data scientist needs a strong theoretical background in SCM along with an ability to apply analysis techniques from a broad variety of quantitative disciplines as well as business disciplines. You will notice that the first six disciplines in Table 1 are quantitative disciplines and the next four are business disciplines. Because data science is applied, tools are needed for application. It is notable that SCM is missing from the list. It is missing from the list because the SCM data scientist must understand both the theory and application of SCM. This is the domain, and the other disciplines are used for application of data to the SCM domain.

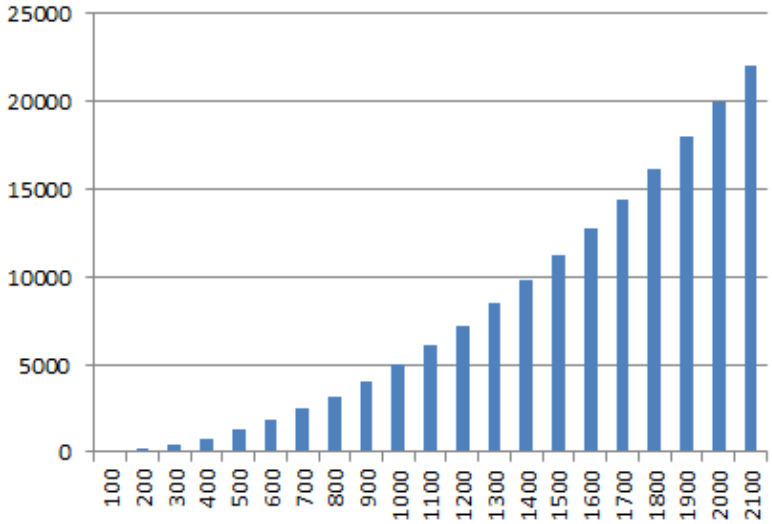
We now propose a definition of SCM data science: *SCM data science is the application of quantitative and qualitative methods from a variety of disciplines in combination with SCM theory to solve relevant SCM problems and predict outcomes, taking into account data quality and availability issues.* We welcome research on this topic and would be pleased to publish it in *Journal of Business Logistics*. Practitioners are looking for answers, and as researchers we should be proposing solutions, frameworks, and answers, all based on theoretically grounded research. We need theoretically based research to verify or reject the ideas in Table 1 and to expand them. At this point, Table 1 is simply conjecture. We invite research that would address which skill sets are needed by SCM data scientists.

Some may object to our claim that SCM theory is needed by the SCM data scientist. However, theory is particularly important now that data and data variety are proliferating. Theory is particularly important for preventing false positives[3]. False positives emerge when relationships between variables

[3] For a very interesting discussion on this topic see Carraway (2012).

are discovered that do not really exist. The problem is that as the number of variables increases—that is, as the use of a-theoretical data mining proliferates—the chances of false positives increases exponentially. In Figure 2, the horizontal axis is the number of variables and the vertical axis is the number of false positives when the probability of a given false positive is 0.01. Theory can help the research or manager avoid spurious decision-making as they avoid falling prey to "apparent" relationships that do not really exist. As Barton and Court (2012) observe, "We have found that … hypothesis-led modeling generates faster outcomes and also roots models in practical data relationships that are more broadly understood by managers." (p. 81)

Figure 2: Relationship between number of variables and number of false positives when the probability of a given false positive is 0.01.



Big data, which will be discussed below, is the source of the explosion of new variables that can be investigated, and therefore the reason why we expect the number of false positives to grow exponentially. Using the *appropriate* logic and/or theory to build models prior to running predictive analytics is a key approach to mitigating the problems associated with false positives. So, again, although SCM data science is applied, it must be based on theory to guard against a proliferation of false positives, which result in wasted time and money. Again, Barton and Court (2012) comment that a "pure data mining approach often leads to an endless search for what the data really say." (p. 81)

Predictive Analytics

Predictive analytics is a subset of data science. Recognition of the uniqueness of predictive analytics illuminates some interesting needs in research as is illustrated by Table 2.

Table 2: Examples of research in predictive analytics

Comparative Discipline	Dimension of Interest	Predictive Analytics Research (EXAMPLES)	
		Relevant	Less Relevant

Statistics	Quantitative	Integrating quantitative and qualitative analysis	Improving Lagrange Multiplier tests for autocorrelation
Forecasting	Predicting the future	Using forecasting techniques for evaluating what would have happened under different circumstances	Deriving generalized estimators of seasonal factors
Optimization	Minimization and maximization	Assessment of the quality of the optimal solution and the ability to implement it versus near optimal solutions	Use of polyhedral functions in linear programming
Discrete event simulation	Quantitative analysis of a system in a stochastic setting	Discrete event simulation in a business process reengineering setting	Random number generation for discrete event simulation
Applied probability	Description of stochastic variables, expected values, and uncertainty	Applied probability along with application anchoring and framing affects from psychology	Asymptotic properties of Gaussian processes
Data mining	Search for patterns and relationships between a large number of variables with lots of data	Data mining preceded by logical and theoretical descriptions of possible relationships and patterns	Gibbs posterior for variable selection in data mining
Analytical mathematical modeling	Precise analysis using artificial and unrealistic assumptions for theorems and proofs	Methods of quickly and inexpensively modeling approximate relationships between variables while still using deductive mathematical methods	Proving inventory theorems that assume known, continuous demand with perfect information

Table 2 examines a sample of disciplines related to predictive analytics, selects a dimension of that discipline, and compares possible research topics and provides an example of a research area that would be more relevant to predictive analytics and an example of a research area that would be less relevant. Table 2 indirectly points to the distinction between predictive analytics and each of these quantitative disciplines. It also provides researchers with possible avenues of research that would be in the realm of predictive analytics.

Importantly, although predictive analytics is related to many long-standing quantitative approaches, it stands as distinct from each. Statistics is quantitative, whereas predictive analytics is

both quantitative and qualitative. Forecasting is about predicting the future, and predictive analytics adds questions regarding what would have happened in the past, given different conditions. Optimization is about finding the minimum or maximum of a function, subject to constraints, whereas predictive analytics also concerns what would characterize a system that was not operating optimally. Analytical modeling is primarily about generating mathematical axioms and then proving lemmas and theorems, whereas predictive analytics attempts to quickly and inexpensively approximate relationships between variables while still using deductive mathematical methods to draw conclusions. These are some examples of the differences in emphasis between predictive analytics and well known quantitative disciplines.

The topics in Table 2 have been examined in part, but additional research in these relevant areas would advance predictive analytics' ability to refine and improve supply chain decision-making. Indeed, the *Journal of Business Logistics* is interested in predictive analytics research that is relevant to logistics and supply chain management. To that end, we propose definitions of logistics and supply chain predictive analytics:

> *Logistics predictive analytics use both quantitative and qualitative methods to estimate the past and future behavior of the flow and storage of inventory, as well as the associated costs and service levels.*
>
> *SCM predictive analytics use both quantitative and qualitative methods to improve supply chain design and competitiveness by estimating past and future levels of integration of business processes among functions or companies, as well as the associated costs and service levels.*

What is defined here as logistics predictive analytics and SCM predictive analytics has already existed in the past, it just lacked a name. The idea is becoming so common that a name helps with communication about the concept. Reading Table 2 with these definitions in mind should provide a guide to appropriate research on logistics or SCM predictive analytics that would be of particular interest at *Journal of Business Logistics*. Barton and Court (2012) highlight the growing value of advanced analytics:

> "Advanced analytics is likely to become a decisive competitive asset in many industries and a core element in companies' efforts to improve performance. It's a mistake to assume that acquiring the right kind of big data is all that matters. Also essential is developing analytics tools that focus on business outcomes …". p. 81

Big Data

Big data is unique because of the volume, variety, and velocity of the data, which today is widely available and much less expensive to access and store (McAfee and Brynjolfsson 2012). Volume can occur in many ways. There are more data because, among other reasons, the data are captured in more detail. For instance, instead of just recording that a unit sold at a particular location, the time it was sold, and the amount of inventory at the time of the sale, is also captured. As another example, many

companies that did not daily sales by location and by stock-keeping-unit (SKU) to make inventory decisions, now do. Moreover, long global supply chains necessitate data capture at multiple points in the supply chain. In addition, there is now a proliferation of consumer sentiment data resulting from Tweets, Likes, and product reviews on websites. Such data must be analyzed and quantified. Software companies that provide algorithms designed to assess text from reviews and Tweets are cropping up in large numbers. Table 3 provides examples of some of the causes of big data.

Table 3: Examples of causes of big data

Type of Data	Volume	Velocity	Variety
Sales	More detail around the sale, including price, quantity, items sold, time of day, date, customer data	From monthly and weekly to daily and hourly	Direct sales, sales of distributors, internet sales, international sales, competitor sales
Consumer	More detail regarding decision and purchasing behavior, including items browsed and bought, frequency, dollar value, timing	From click through to card usage.	Face profiling data for shopper identification and emotion detection; eye tracking data; customer sentiment about products purchased based on "Likes," "Tweets," and product reviews
Inventory	Perpetual inventory at more locations, at a more disaggregate level (e.g., style/color/size)	From monthly updates to hourly updates	Inventory in warehouses, stores, internet stores, and a wide variety of vendors online
Location & Time	Sensor data to detect location in store, including misplaced inventory, in distribution center (picking, racks, staging, etc.), in transportation unit	Frequent updates for new location and movement	Not only where it is, but what is close to it, who moved it, its path to get there, and its predicted path forward; location positions that are time stamped from mobile devices

Table 4 provides examples of potential applications of big data within logistics and SCM practice. Each Column of Table 4 represents a key managerial component of business logistics and each row

represents a different category of user of logistics. This is not intended to be an exhaustive list of components of logistics nor of users of logistics.

Table 4: Examples of potential applications of big data in logistics

User	Forecasting	Inventory Management	Transportation Management	Human Resources
Carrier	Time of delivery, factoring in weather, driver characteristics, time of day and date	Real time capacity availability	Optimal routing, taking into account weather, traffic congestion, and driver characteristics	Reduction of driver turnover, driver assignment, using sentiment data analysis
Manufacturer	Early response to extremely negative or positive customer sentiment	Reduction in shrink, ECR, quick response, VMI	Improved notification of delivery time, and availability; surveillance data for improved yard management	More effective monitoring of productivity; medical sensors for safety of labor in factories
Retailer	Customer sentiment data and use of mobile devices in stores	Improvement in perpetual inventory system accuracy	Linking local traffic congestion and weather to store traffic	Reduction in labor due to reduction in misplaced inventory

Table 5 provides examples of research questions based upon management areas within logistics and supply chain management with reference to various sources of big data.

Table 5: Examples of research questions

Type of Data	Inventory Management	Transportation Management	Customer & Supplier Relationship Management
Sales	How can sales data be used with detailed	How can more current sales data be used to	How can more granular sales data from the

	customer data to improve inventory management either in terms of forecasting or treating some inventory as "committed" based on specific shoppers requirements?	re-direct shipments in transit? How can sales data, integrated with detailed customer data, be used for more efficient and effective merge-in-transit operations?	wide variety of sources that exist be used to improve visibility on the one hand and trust on the other, between trading partners?
Consumer	How can face profiling data for shopper identification, emotion detection and eye tracking data can be used to determine which items to carry and stock at particular shelf locations?	How can delivery preferences captured in online purchases be used to manage transportation mode and carrier selection decisions?	How can customer sentiment about products purchased based on "Likes," "Tweets," and product reviews be used to collaborate on forecasts?
Location & Time	How can sensor data used to detect location in store, be used to improve inventory management, including departmental merchandising decisions?	How can sensor data in the distribution center be used to anticipate transportation requirements?	How can location and time-stamp data of shoppers be used for collaborative assortment and merchandising decisions?

Back to Basics

Finally, using management theory as a lens, we provide a few examples of research questions that are relevant to supply chain management.

Table 6: Examples of big data research questions that are relevant to SCM, stemming from management theory

Theory	**Research Question**
Transaction Cost Economics	How does the existence of big data affect the reduction of internal transaction costs vis-à-vis external transaction costs and how is this affecting the size of logistics organizations and the structure of supply chains?
Resource Based View	Can SCM data science be developed as a resource that is valuable, rare, in-imitable, and non-substitutable?
Contingency Theory	How can big data and SCM data science used by logistics managers to

	meet internal needs and adjust to changes in the supply chain environment?
Resource Dependence Theory	How does the ability to use big data for supply chain management decisions affect a firm's power in comparison to its suppliers or customers?
Agency Theory	How does the proliferation of big data affect the agency costs associated with the use of third party logistics?
Institutional Theory	How do differences in freedom of information between countries affect firms' operating under these different institutions in terms of their abilities to leverage big data in the supply chain?

Demand for DPB Professionals

We believe that there is increasing demand for professionals with competencies in DPB. As an example of overall interest, we show figures of the increasing numbers of Google searches for these terms. Figure 3 show a graph of numbers of Google searches[4] since 2004 for various relevant terms. The scales on the y-axes are relative with 100 representing the peak number of searches.

Figure 3: Panel A: Google Searches for "Data Science" and "Data Scientist" since 2004

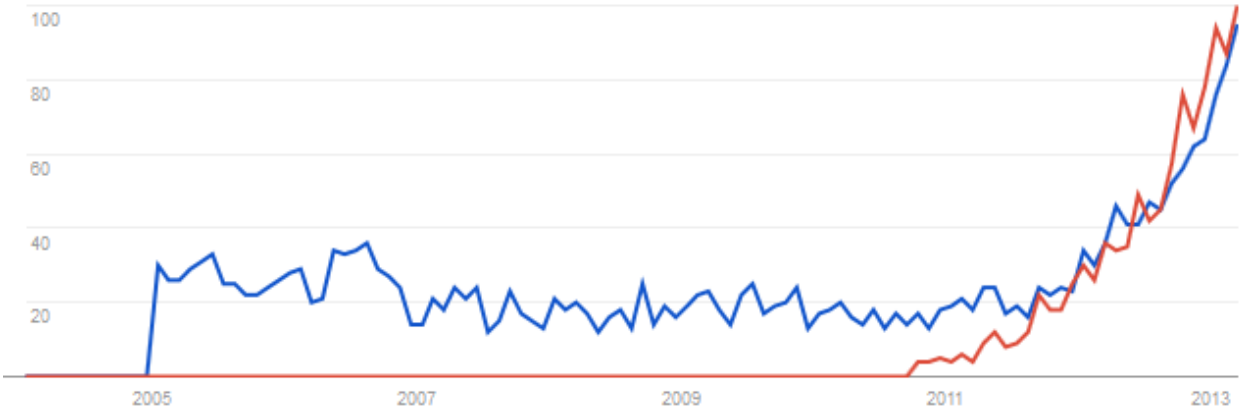


Figure 3: Panel B: Google Searches for "Predictive Analytics" since 2004

[4] http://www.google.com/trends/explore#q=%22data%20science%22%2C%20%22data%20scientist%22&cmpt=q referenced March 22, 2013.

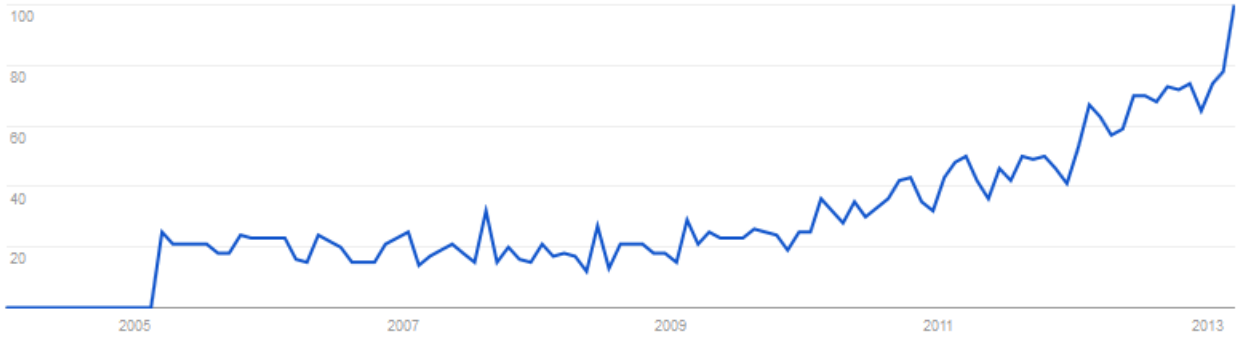
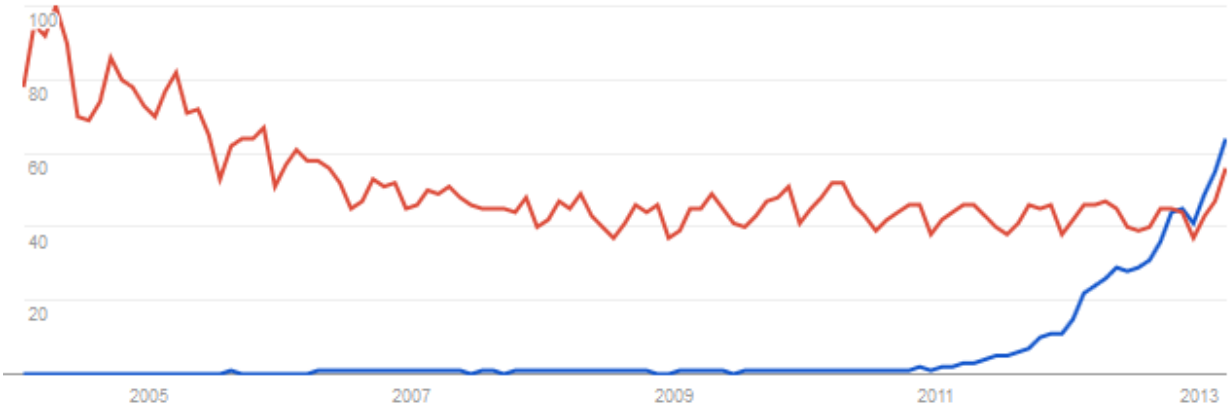


Figure 3: Panel C: Google Searches for "Big Data" and "Supply Chain Management" since 2004



As seen in Figure 1, there were virtually no searches for "Data Science" until 2005 and none for "Data Scientist" until 2011. We believe the terms are catching on because of the increased need for people with skill sets that can deal with big data. Figure 2 shows a graph of searches[5] for "Predictive Analytics." Like "Data Science," searches for "Predictive Analytics" essentially began in 2005 with significant growth after 2009. Figure 3 shows a graph of searches[6] for "Big Data," with "Supply Chain Management" as a reference. As you can see, this is the first year "Big Data" had more searches than "Supply Chain Management." We do not think big data will become more important than supply chain management; it is shown here just to illustrate how popular the phrase is becoming.

Clearly, Figure 3 shows that interest in DPB is growing exponentially. We believe that the phenomena underlying this trend create a number of challenges and opportunities for our discipline. We have only begun to explore the possibilities and we need to more creatively ask informed decisions about how big data can improve supply chain design, relationship development, improve customer service systems, and manage day-to-day value-added operations.

[5] http://www.google.com/trends/explore#q=%22predictive%20analytics%22&cmpt=q referenced March 22, 2013.

[6] http://www.google.com/trends/explore#q=%22big%20data%22%2C%20%22supply%20chain%20management%22&cmpt=q referenced March 22, 2013.

CONCLUDING DISCUSSION

Although "Big data" has becomes a contemporary buzzword, it has significant implications in our discipline, and presents an opportunity and a challenge to our approach of research and teaching. We can easily see how data science and predictive analytics apply to SCM but sometimes find it more difficult to see the direct connection of big data to SCM. Therefore, we would like to see research published in *Journal of Business Logistics* that brings clarity to the relevance of big data, and DPB in general within the supply chain domain. Here is how you can participate:

1. Submit manuscripts dealing with DPB
2. Submit Forward Thinking articles on DPB
3. Send us a proposal for a Special Topics Forum on DPB
4. Design a Thought Leader Series on DPB
5. Start a new research project on DPB, using your existing research skills and domain knowledge
6. If you have ideas about how we can promote research on DPB outside of these categories, please email us and let us know your thoughts

This edition of *Journal of Business Logistics*, Volume 34, Issue 2, represents the mid-point of our five-year commitment as co-editors-in-chief of the *Journal*. We have been diligent in this commitment to serve not only as stewards and administrators of the journal but also to span boundaries to provide leadership and direction for research in our discipline. We believe the intersections of our discipline with data science, predictive analytics, and big data create significant challenges for educating future supply chain leaders. Yet, they also provide opportunities for research to advance knowledge in our discipline in a way that is both rigorous and relevant. Indeed, we believe it is a real doozy for knowledge creation and dissemination in supply chain management.

Figure 4: Word cloud of this Editorial

ACKNOWLEDGMENTS

We thank the following individuals for their helpful comments, edits and input, on earlier drafts of this manuscript: Yao (Henry) Jin, John Saldana, Travis Tokar, Christopher Vincent, Xiang Wan, and Brent Williams. Their input resulted in a significant improvement to the manuscript.

REFERENCES

Barton, D. and Court, D. 2012. “Making Advanced Analytics Work for You,” *Harvard Business Review*, October: 79-83.

Carraway, R. 2012. “Big Data, Small Bets,” *Forbes*, December 12, http://www.forbes.com/sites/darden/2012/12/13/big-data-small-bets/.

Chen, H., Chiang, R., Storey, V. 2012. “Business Intelligence and Analytics: From Big Data to Big Impact,” *MIS Quarterly*, 36(4): 1165-1188.

Davenport, T., and Patil, D. 2012. “Data Scientist: The Sexiest Job of the 21st Century,” *Harvard Business Review*, October: 70-76.

Dumbill, E, Liddy, E., Stanton, J., Mueller, K., and Farnham, S. 2013. “Educating the Next Generation of Data Scientists,” *Big Data*, March 1(1):21-27.

McAfee, A., and Brynjolfsson, E. 2012. “Big Data: The Management Revolution,” *Harvard Business Review*, October: 60-68.

Provost, F., and Fawcett, T. 2013. “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” *Big Data*, March 1(1):51-59.