CONSIDERING DATA QUALITY FOR DATA SCIENCE, PREDICTIVE ANALYTICS, AND BIG DATA APPLICATIONS IN SUPPLY CHAIN MANAGEMENT

Benjamin T. Hazen, Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 919-722-8218, benjamin.hazen@us.af.mil
Christopher A. Boone, Department of Marketing and Logistics, College of Business Administration, Georgia Southern University, Statesboro, GA 30458, 912-478-6625 caboone@georgiasouthern.edu
Jeremy D. Ezell, Department of Aviation and Supply Chain Management, Harbert College of Business, Auburn University, AL 36849, 334-844-8299, jde0009@tigermail.auburn.edu
L. Allison Jones-Farmer, Department of Aviation and Supply Chain Management, Harbert College of Business, Auburn University, AL 36849, 334-844-8299, ajones@auburn.edu

ABSTRACT

Today's supply chain professionals are inundated with data, motivating new ways of thinking about how data are produced, organized, and analyzed. This has provided an impetus for organizations to adopt and perfect data science, predictive analytics, and big data (DPB) functions in order to enhance supply chain processes and, ultimately, performance. However, the decisions made as a result of DPB are only as good as the data on which they are based. In this presentation, we introduce the data quality problem in the context of DPB for supply chain management (SCM) and highlight interdisciplinary research topics based on complementary theory.

THEORY-BASED RESEARCH OPPORTUNITIES

While there are surely several theories that can be used as a basis to study emerging problems regarding data quality in the context of DPB and SCM, as a starting point, we frame some example research questions in the context of the knowledge-based view (KBV), systems theory, and the organizational information processing view (OIPV).

Knowledge-Based View

The resource-based view (RBV) suggests that organizations use rare, valuable, and inimitable resources to create competitive advantage and is often used to frame research in the SCM domain. As an extension of RBV, the KBV considers knowledge as one such resource. Thus, it might be especially important to consider data quality when determining the degree to which DPB can be utilized as a knowledge resource that has a potential to enhance competitive advantage. Considering the KBV, we suggest the following research questions:

- Does data quality mediate the relationship between DPB activities and measures of supply chain performance? If so, what is the nature of this mediation?
- Is data quality a direct antecedent to DPB usage in supply chain applications? When examining the effects of DPB, is data quality a direct antecedent to supply chain performance?

Systems Theory

Systems theory is another commonly used theory in the SCM literature and suggests that organizations are systems that are open and interact with their environment, and thus are continually evolving. From this perspective, it is easy to envision a supply chain as a system of connected nodes that are interacting with and relying on inputs from the external environment. Similarly, information systems that support DPB can be viewed as SCM sub-systems operating within a real-world feedback control system. Orr suggests that there must be a mechanism to synchronize system data with changes in the environment. As such, measuring and controlling data quality might be especially important from a systems theory perspective when investigating the impact of DPB on SCM performance. Considering systems theory, we suggest the following research questions:

- How can organizations integrate data quality and control initiatives into their existing or emerging supply chain DPB programs?
- What are the costs of data quality to specific supply chain DPB initiatives? What are the costs of maintaining "acceptable" data quality, and do these costs result in an adequate return on investment? What is "acceptable" data quality for supply chain operations?
- How does one firm's data quality affect the DPB efforts of partner firms within the supply chain?
- Considering the boundary-spanning role of logistics operations, how does data quality affect firm processes outside of logistics operations?

Organizational Information Processing View

The OIPV suggests that organizations are imperfect decision-making systems due to incomplete information. The theory considers three primary components: information processing needs, information processing capabilities, and the fit between needs and capabilities. Information processing needs indicate the information required by the organization to enable effective decision-making, whereas information processing capabilities indicate the organization's actual capacity to utilize and structure information to support decision-making. Fit is the degree to which a firm's information processing capabilities satisfy its information processing needs. It is via fit that firms can enhance decision-making capabilities and, subsequently, performance. As such, OIPV can be a useful lens through which to examine how efforts to enhance data quality might lead to increased levels of performance realized via DPB. Considering the OIPV, we suggest the following research questions:

- Considering the importance of fit, does data quality play a moderating role in relationships between various independent variables and logistics performance garnered from DPB?
- Does data quality affect the relationship between supply chain information processing needs and capabilities?

CONCLUSION

As alluded to herein, the study of DPB in general and the data quality problem in particular requires interdisciplinary collaboration in order to advance. For instance, information systems experts are needed to provide insight into how data are collected, stored, processed, and retrieved. SCM domain experts are needed to ensure that the right problems are being addressed and the analysis being performed and results derived thereof are relevant. Additionally, emerging data quality research suggests the need for statistical and analytical experts who are knowledgeable in methods required to measure, monitor, and control data quality. Working together, scholars from these and other disciplines can employ the right technologies and techniques to solve the right problems.