What is Big Data? : Findings from a systematic literature review, Hace theorem and a case study

Mohamed Hicham Salah Eddine
Research team AMIPS, Ecole Mohammadia d'Ingénieurs,
Mohammed V University of Rabat, Morocco
medhichamsalaheddine@research.emi.ac.ma

Tarik Saikouk
International logistics and supply chain department
International University of Rabat, Morocco
tarik.saikouk@uir.ac.ma
Rabat, Morocco

Abdelaziz Berrado
Research team AMIPS, Ecole Mohammadia d'Ingénieurs,
Mohammed V University of Rabat, Morocco
berrado@emi.ac.ma

Abstract

Big data" has recently become the focus of academic and corporate research. The literature identifies 'big data' as the 'next big thing in innovation. Drawing on a systematic review and a case study findings, this paper presents an interpretive framework that analyses the definitional perspectives and the applications of big data. We used the Hace theorem to understand the concept and a specific classification framework to have a better understanding of its role in capturing business value. This was further confirmed through a case study in the debt collection field where big data was used to find insolvent debtors.

Keywords

Big Data, Hace theorem, Debt Collection

Introduction

Every day 3 billion kilobytes of data are produced. Today 90 percent of the data in the web where created within the last two years. (Tamhane, sayyade, 2015), Big data has the potential to revolutionize the art of management (wamba, 2013), Due to its high operational and strategic potential, notably in generating business value, "big data" has recently become the focus of academic and corporate investigation. The extant literature identifies 'big data' as the 'next big thing in innovation' (Gobble, 2013); "the fourth paradigm of science" (p. 34) Strawn (2012); ''the next frontier for innovation, competition, and productivity''(p. 1) (Manyika et al., 2011); the next "management revolution" (p. 3)(McAfee & Brynjolfsson, 2012); and a revolution in science and technology" (p. 4) (Ann Keller, Koonin, & Shipp, 2012), etc. In addition, 'big data' has the capability of transforming the decision making process by allowing enhanced visibility of firm operations and improved performance measurement mechanisms (McAfee & Brynjolfsson, 2012). The aim of this work is to presents an interpretive framework that analyses the definitional perspectives and the applications of big data then to introduce a case study of N Square Group where Big Data where used to find debt collector in Morocco.

Some studies estimate an increase in annually created, replicated and consumed data from around 1,200 Exabyte's in 2010 to 40,000 in 2020 (Gantz and Reinsel, 2012). In some industries big data has led to the creation of entirely new business models. In the retail sector, big data expedites the analysis of in-store purchasing behavior in near real-time to adjust merchandise, stock levels and prices (Hagen et al., 2013). Although companies relying on data – such as insurance companies – is not a new concept, it was only recently that companies began to try to make use of other data sources such as social media, smartphones or sensors, and new technologies to open to new perspectives.

Despite the excitement and recent interest in 'big data', due to its high operational and strategic potential, little is known about what encompasses the concept, very few empirical studies have been conducted to assess it's real potential.

In the first phase of the study, a comprehensive review of articles dealing with 'big data' related topics was adopted based on a similar approach used by Ngai and Wat (2002) in electronic commerce, (E. W. T. Ngai, Xiu, & Chau, 2009),in CRM and data mining (E. W. T. Ngai, Moon, Riggins, & Yi, 2008), (Samuel Fosso Wamba, Abhijith Anand, & Lemuria Carter, 2013),RFID related topics and (Wamba, Akter, Edwards, Chopin, Gnanzou,2017)in big data literature review.

The approach entails three key characteristics: (i) the development of a classification framework; (ii) conduct the literature review and (iii) realize the classification of relevant journal articles. As recommended by (W. T. Ngai & Wat, 2002).

1. Literature review

The starting point for the review was to identify the operations management journals to be included in the study. In order to include a range of disciplines within operations management and to report only on research of a high quality, three and four star operations management journals were selected from the Association of Business Schools journal ranking guide (www. charteredabs.org). Within the business and management field including economics, there are a small number of grade 4 journals that are recognized world-wide as exemplars of excellence. Their high status is acknowledged by their inclusion in a number of well-regarded international journal quality lists. The Guide normally rates a journal 4* if they are rated in the highest category by at least three out of the five non-university based listings. A search within the timeframe ranging from 2006 to 2017 was considered to be representative of the period covering the apotheosis of 'big data', with a comprehensive search using the descriptor, "big data" conducted within the following databases: ABI/Inform Complete, Academic Search Complete, Business Source Complete, Elsevier, Emerald, IEEE Xplore, Science Direct, and Taylor & Francis. The study was realized within the top journals of the business and management field: Journal of Operations Management, International Journal of Operations and Production Management, Production and Operations Management, Computers in Industry, IEEE Transactions on Engineering Management, International Journal of Production Economics, International Journal of Production Research, Journal of Scheduling, Journal of Supply Chain Management, Manufacturing and Service Operations Management, Production Planning and Control, Supply Chain Management: An International Journal.

Business and management field literatures provide important insights into the implementation, adoption and use of big data in the supply chain management field, as well as its business value (wamba,(2015), Tambe,(2014), Hartmann,(2016))− which may be considered as a key facilitator or enabler of 'big data' when exploring its potential by managers.
As suggested by the approaches of Thorpe et al. (2005) and Macpherson and Holt (2007), the titles, abstracts and key words were searched for the term of "big data", and "big data supply chain management" A total of 1718 studies were identified and the number of papers extracted for each journal are listed in Table 1.

Table 1: Characteristics of the reviewed papers

Journal Title	**RANK**	**Databases used**	Big data and SC M	BIG DATA
Journal of Operations Management	4*	science direct	0	43
International Journal of Operations and Production Management	4	emerald	47	76
Production and Operations Management	4	Wiley	26	38
Computers in Industry	3	science direct	39	174
IEEE Transactions on Engineering Management	3	IEEE	4	51
International Journal of Production Economics	3	science direct	150	433
International Journal of Production Research	3	Taylor and Francis online	168	664
Journal of Scheduling	3	Springer	47	91
Journal of Supply Chain Management	3	Wiley	24	35
Manufacturing and Service Operations Management	3	Informs	13	23
Production Planning and Control	3	Taylor and Francis online	93	141
Supply Chain Management: An International Journal	3	emerald	37	0
		total	648	1718

To filter out any irrelevant references to supply chain and big data, following Thorpe et al. (2005) and Macpherson and Holt (2007), the second step was to identify the search term(s) and to search the journals identified above. Consistent with the approaches of Thorpe et al. (2005) and Macpherson and Holt (2007), the titles, abstracts and key words were searched for the exact term of "big data". However, due to the size of the field, we added the term "big data; supply chain" as a total of 648 studies were returned. The number of papers extracted for each journal are listed in Table 1.

The third stage was to filter out any irrelevant references to big data and supply chain, following Thorpe et al. (2005) and Macpherson and Holt (2007), resulting in a total of 283 paper being removed. 179 articles were removed as Big data was referenced either in the biography/references, as part as a editorial comment or in a literature review. 102 papers were then removed as their reference to big data as we judged that it was unrelated to the core argument of the paper, for example (Craighead et al., 2009). Finally, 58 papers were removed as big data was used indirectly, such as the use of Big data in biocuration (Howe et al., 2008) .

Finally, 26 papers were retained out of a total of 1189 papers, as they were deemed relevant understand the concept of big data and its potential relationship with supply chain management.

Table 2: Papers retained through the literature review

Journal Title	**RANK**	**Databases used**	Accepted articles
Journal of Operations Management	4*	science direct	0
International Journal of Operations and Production Management	4	emerald	4
Production and Operations Management	4	Wiley	0
Computers in Industry	3	science direct	2
IEEE Transactions on Engineering Management	3	IEEE	0
International Journal of Production Economics	3	science direct	11
International Journal of Production Research	3	Taylor and Francis online	9
Journal of Scheduling	3	Springer	0
Journal of Supply Chain Management	3	Wiley	0
Manufacturing and Service Operations Management	3	Informs	0
Production Planning and Control	3	Taylor and Francis online	0
Supply Chain Management: An International Journal	3	emerald	0
		total	26

2. Big Data Characteristics: HACE Theorem

2.1. Definition of big data

Why should academics and practitioners be interested in understanding the big data? The simple answer to this critical question is because big data has the potential to transform the entire business process (wamba,2015). Wamba (2015) define "big data" as a "as a holistic approach to manage, process and analyze the data in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages, it's all the data captured from sensors, posts to social media

sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals, (ibm,2012b), Johnson (2012) define it as an extremely large sets of data related to consumer behavior, social network posts, geotagging, sensor outputs. For Manyika et al (2011) big data refers to a datasets with a size that is beyond the ability of typical database software tools to capture, store, manage, and analyze. Gardner (2012) refer to « big data » as a high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. One of the main objective of this study is to clearly understand Big Data as a whole.

Imagine that a number of blind men are trying to size up a giant elephant, which will be the Big Data in this context. The goal of each blind man is to extract conclusion of the elephant according to the part of information he collects during the procedure (Deepak and Sayyade(2015), Because each person's view is limited to his local region, it is not surprising that the blind men will each conclude independently that the elephant ―feels like a rope, a hose, or a wall, depending on the region each of them is limited to. To make the problem even more complicated, let us assume that the elephant is growing rapidly and its pose changes constantly, and each blind man may have his own (possible unreliable and inaccurate) information sources that tell him about biased knowledge about the elephant (e.g., one blind man may exchange his feeling about the camel with another blind man, where the exchanged knowledge is inherently biased.by exploring « big data »,we are aggregating different types of information from different sources to draw the best approximate picture of the reality. The HACE theorem as suggested by Kalaivani (2015) suggest that the key for big data characteristics are :

- Huge with heterogeneous and diverse data sources: important volume with diverse dimensionalities,
- Decentralized control: Autonomous data sources, so that each data sources can freely generate data,
- Complex data and knowledge associations: the impressive diversity of data imply that it require a « big mind»to consolidate data for maximum values.

The study was conducted with a group of doctoral students working on either supply chain management or data mining related research topics and as proposed with Hace theorem concept, different brainstorming sessions were held to discuss the definition of the concept.
Big Data has three main characteristics: the data itself, the analytics of the data, and the results of the analytics. Then there are the products and services that can be wrapped around one or all of these Big Data elements (Idc,2013), Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data, These characteristic

2.2. Distribution of articles by the type of value creation from 'big data'

First, we can notice that many of the publications covered more than one type of value creation from 'big data'. Clearly, the vast majority of the publications are in 'Enabling experimentation to discover needs, expose variability, and improve performances' (13 articles or 52% of all publications), performance improvement is at the core of the current hype around 'big data'. Followed by 'Innovating new business models, products, and service' and ''Segmenting populations to customize actions' with respectively 4 articles (or 16% of all publication) and 4 articles (or 16% of all publications). Finally, we have 'Replacing/supporting human decision making with automated algorithms' with 2 articles (or 2% of all publications), and 'Creating transparency' with 2 articles or 8% of all publications.

Table 3: Classification of papers according to the type of value creation

Dimension	Reference	#	%
creating transparency	(huang,2015);(fantazzini,2015)	2	8%
Enabling experimentation to discover needs, expose variability, and improve performance	(kwon,2014);(kartman,2016);(koche,2017);(tan, 2015);(wamba2015);(akter,2015);(brusset,2016) ; (wang,2016);(jifan,2017);(chan,2015);(hofman, 2015);(li,2015); (pang,2017)	13	52%
Segmenting populations to customize actions	(kaulio,2016);(sheng,2017);(Fantozzini,2015);(wang,2016)	4	16%
Replacing/supporting human decision making with automated algorithms	(pang,2017);(kaur,2015)	2	8%
Innovating new business models, products, and service	(dutta,2015);(fantozzini,2015);(zhong,2017) ;(o presnik,2015)	4	16%

2.3. **Distribution of articles by the type of issues related to 'big data'- enabled business value**

Not surprisingly, the vast majority of articles (44%) have issues related to access to data, many scholars attempt to understand how to access data from various sources and see the information mining as the one of the critical leverage to improve performance.

Table 4: Classification of papers according to the type of issues related to 'big data'

Issues	**References**	**#**	**%**
Data policies	(wamba,2015)	1	4%
Technology and techniques	(kaur,2015);(kache,2017),(fantozzini2015),(zhong,2017),(pang,2017)	5	19%
Organizational change and talent	(mandar,2016);(hartman,2016);(sheng,2017);(dutta,2015);(opresnik,2015);(wang,2016)	6	22%
Access to data	(kwon,2014);(huang,2015);(kache,2017);(fortazzini,2015);(wamba,2015);(brusset,2016);(ren,2017);(chan,2015);(wang,2016);(hofmann,2015);(li,2015)	12	44%
Industry structure	(mandar,2016);(akter,2016);(li,2015)	3	11%

2.4. Distribution of article by industry

The highest number of published articles is in 'manufacturing' industry (9articles or 39% of all publications), which may highlight the fact that 'big data' is increasingly viewed as a practical concept that can improve the performance, as we could not classify 7 (30%of all publications) in a specific industry.

Table 5: Classification of reviewed papers by industry

Industry	**References**	**#**	**%**
manufacturing	(kwon,2014);(kaur,2015);(huang,2015);(dutta,2015);(akter,2016);(wang,2016);(hofmann,2015);(mandar,2016);(zhong,2017)	9	39%
other	(hartman,2016);(kache,2017);(sheng,2017);(brusset,2016);(wang,2016);(Ren,2017);(chan,2015)	7	30%
healthcare	(wamba,2015)	1	4%
retail	(hofmann,2015);(li,2015)	2	9%
service	(opresnik,2015)	1	4%
technology	(fontazzini,2015),(tan,2015);(pang,2017)	3	13%

3. Insights and lessons learned from the case study

In the following sections, we will present key insights and lessons learned from the in-depth analysis of the longitudinal case study of "N Square- Group" which is currently using 'big data' for improved operations of debt collecting.

The better management of collecting operations required the integration of multiple sources of data (structured and unstructured) across multiple sources; the combination of these data with historical information for better case management. In the case of N Square group, the company has developed a range of IT capabilities over time. For example, N square group has a bi-directional direct communication between its IT system and the BODACC "Official Bulletin of Civil and Commercial Advertisements". The same capabilities allow N square group to share debt collector information with its principal clients. The organization has been aggressively using cutting edge tools and technologies such as telephony, location, real time, smart phone application, communications, mapping tools, and specific IT system in order to provide improved capabilities to independent field workers during collecting operations. A typical response operation would involve multiple information gathering, processing and dissemination technologies to find insolvent debtor.

Debt collecting operations are directed from a command control center at the company headquarters. A dashboard is a key resource employed by teams coordinating responses to specific portfolio, specific clients and specific business units. Among the data elements displayed on the dashboard is real-time data coming from different sources such as Facebook, Page Jaune, Twitter, Charikati … through a direct link to display potential information that may affect the collecting process. The relevant information is then routed via various channels, including the company IT system P4C, all the independent debt collector field worker's Smartphones to the clients' specific platform.

Integrating all these sources of data for improved Debt collecting process is achieved through a single IT instance as their shared platform by all the principal clients (banks, loan organization and industries), the 'backbone' of the IT infrastructure that allows a connection between the agencies and both the various data sources and existing legacy systems containing structured data, unstructured data and historical data (e.g., records of all publicity of deeds registered in the register of commerce and companies) related to write-offs, transfer of shares.

One of the key lessons learned from this case study is the importance of the active engagement of the

team that implemented the new IT-enabled 'big data' infrastructure in collaboration with the employees – especially the independent debt collector – during the whole project. As key stakeholders in the project, they were given the opportunity to contribute as to understand, based on their different past experiences, what the use of different information on the debt collecting process and how to give a meaning to them. The active engagement of the banks and loan organization was mandatory first because of the information of their insolvent debtor and the funding for the whole project witch cost 3 years of labor.

Key insights from the in-depth case study indicate that creating and capturing business value from 'big data' can allow a real-time access and sharing of information across all the supply chain not only for improved decision-making but to literally enhance the collecting rate. For example, having real-time information on 'who' and 'where' is allowing not only the realignment and use of the right resource across the country, also informing strategic decision about where and when to program a large scale house visits. Another key benefit realized from 'big data' by the N square group is the improvement of intra- and inter-organizational transparency and accountability, which represent major issues in the business environment. Moreover, the ability of this company to handle and support data from various sources and formats (structured and unstructured), as well as to push 'intelligence' from these data to various channels so as to support emergency operation on the field, is relatively new and can be a critical success factor in this process of creating and capturing business value from 'big data' in the debt collection field, and it needs continues improvement work.

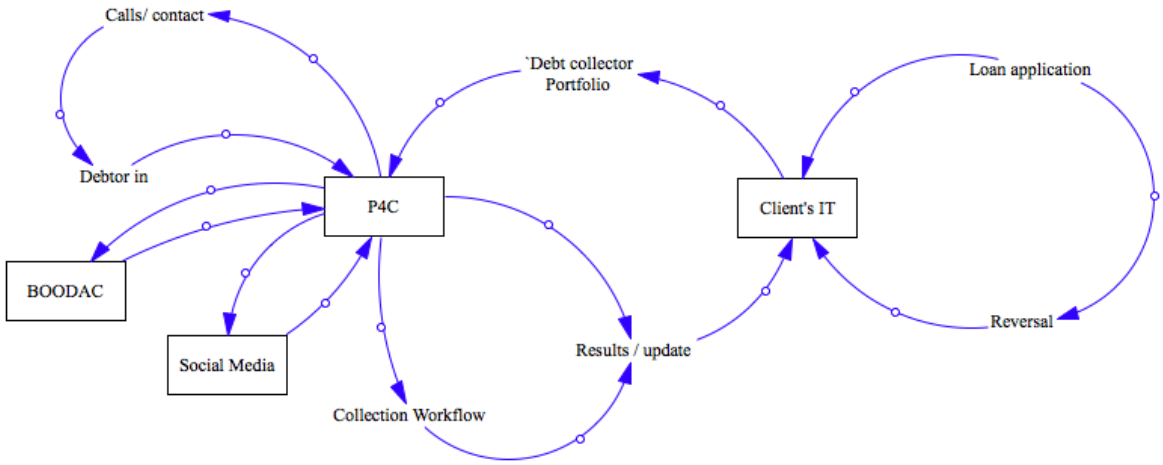

Figure 1: Debt collector case study

4. Conclusions and future research

To date, most of the studies undertaken on Big Data have focused on a wide range of issues and are generally qualitative in nature. This research applied a qualitative approach not only to understand the concept, but also the practical use of it. However, these assumption need to be tested by large-scale empirical studies. As with any relatively new research area, the conceptual model presented in this paper is just one of the possible views. As such, it is an obvious limitation. The idea of establishing a meaningful linkage between the 5'v and big data was a central concept for this research and in order to justify the need for it, one needs to have an understanding and clear definition of the phenomenon of big data. The increased risks that are the result of complex and geographically disperse global supply chains necessitates that companies gain a better theoretical understanding of this emerging critical topic in order to effectively manage in this business environment and major companies need to invest more and more in

this field.

References:

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Brusset, X. (2016). Does supply chain visibility enhance agility? *International Journal of Production Economics*, *171*, 46–59.

Chan, H. K., Lacka, E., Yee, R. W. Y., & Lim, M. K. (2015). The role of social media data in operations and production management. *International Journal of Production Research*, *7543*(July 2015), 1–10.

Dutta, D., & Bose, I. (2015). Managing a big data project: The case of Ramco cements limited. *International Journal of Production Economics*, *165*, 293–306.

Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, *170*, 97–135. https://doi.org/10.1016/j.ijpe.2015.09.010

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

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, *36*(10), 1382–1406.

Hofmann, E. (2017). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *55*(17), 5108–5126.

Huang, Y.-Y., & Handfield, R. B. (2015). Measuring the benefits of ERP on supply management maturity model: a "big data" method. *International Journal of Operations & Production Management*, *35*(1), 2–25.

Ji-fan Ren, S., Fosso Wamba, S., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011–5026.

Kache, F., & 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*, *37*(1), 10–36.

Kaur, N., & Sood, S. K. (2015). Cognitive decision making in smart industry. *Computers in Industry*, *74*, 151–161.

Tim Baines Ali Ziaee Bigdeli Oscar F. Bustinza Victor Guang Shi James Baldwin Keith Ridgway , (2017)," *Servitization: revisiting the state-of-the-art and research priorities* ", International Journal of Operations & Production Management, Vol. 37 Iss 2 pp. 256 - 278

Kwon, K., Kang, D., Yoon, Y., Sohn, J.-S., & Chung, I.-J. (2014). A real time process management system using RFID data mining. *Computers in Industry*, *65*(4), 721–732.

Li, D., & Wang, X. (2017). Dynamic supply chain decisions based on networked sensor data: an application in the chilled food retail chain. *International Journal of Production Research*, *55*(17), 5127–5141. https://doi.org/10.1080/00207543.2015.1047976

Mandar Dabhilkar Seyoum Eshetu Birkie Matti Kaulio , (2016),"Supply-side resilience as practice bundles: A critical incident study", International Journal of Operations & Production Management, Vol. 36 Iss 8 pp.

Opresnik, D., & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics*, *165*, 174–184.

Pang, K.-W., & Chan, H.-L. (2017). Data mining-based algorithm for storage location assignment in a randomised warehouse. *International Journal of Production Research*, *55*(14), 4035–4052.

Prasanna Tambe (2014) Big Data Investment, Skills, and Firm Value. Management Science 60(6):1452-1469.

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, *191*(November 2016), 97–112.

Tan, K. H., Zhan, Y. Z., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, *165*, 223–233.

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

Wang, J., & Zhang, J. (2016). Big data analytics for forecasting cycle time in semiconductor wafer fabrication system. *International Journal of Production Research*, *54*(23), 7231–7244.

Zhong, R. Y., Xu, C., Chen, C., & Huang, G. Q. (2017). Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors. *International Journal of Production Research*, *55*(9), 2610–2621.

Biographies

Mohamed Hicham Salah Eddine is holder of a master degree in international logistics and supply chain management from the international university of rabat and a phd student at Ecole mohammadia d'ingénieur, his research focuses on big data, supply chain resilience and it's dynamic complexity, specially the debt collecting field, he fonded in 2017 "N Square group", a start up that use advenced technologie to help loan compagnies to search find and manage insolvent debtor.

Tarik Saikouk is holder of an engineering degree in industrial systems engineering in 2009 from the Université de Technologies de Troyes and a PhD in management sciences in Supply Chain Management in 2013 at the University of Grenoble.Dr. Tarik SAIKOUK is currently a teacher-researcher at the Rabat Business School (RBS) at the International University of Rabat and also visiting teacher at the ESC Rennes in France. He is also responsible for the International Logistics and Supply Chain Management Master at the RBS. Also, head of the IL & SCM Master in Initial Training and Head of the Executive Master in Global Supply Chain Ecosystem in partnership with SNTL in Morocco and founding member of the Moroccan association of the aeronautical supply chain. His research, mainly empirical, focuses on the dynamics and complexity of the supply chain, strategic behaviors within the supply chain, the supply chain maturity of companies, technologies of traceability and continuous improvement process (Lean management & Theory of Constraints).

Abdelaziz Berrado, Ph.D, is Department Chair and Associate Professor of Industrial Engineering in EMI School of Engineering at Mohamed V University in Rabat. He holds a Ph.D. degree in Decision Systems and Industrial Engineering from the Ira A. Fulton School of Engineering at Arizona State University. His research, teaching and consulting interests are in the areas of Big Data Analytics, Industrial Statistics, Operations and Supply Chain Modelling, Planning and Control with applications in healthcare, education and other industries. He focuses on developing frameworks, methods and tools for systems" diagnostics, optimization and control with the aim of operational excellence. He published several papers in research journals and conferences with local and international funding. He reviews for many journals and is a member of INFORMS, IEOM and IEEE. In addition to academic work, he interacts continuously with different Industries in the areas of Machine Learning, Quality Engineering and Supply Chain Management. He was also a senior engineer at Intel.