A Risk and Benefits Behavioral Model to Assess Intentions to Adopt Big Data

José Esteves[1] and José Curto[2]

[1]E Business School, Spain [2]UOC, Spain

Email: jose.esteves@ie.edu jcurtod@uoc.edu

ABSTRACT: Everyday a constant stream of data is generated as a result of social interactions, Internet of things, e-commerce and other business processes. This vast amount of data should be collected, stored, transformed, monitored and analyzed in a relatively brief period of time. Reason behind is data may contain the answer to business insights and new ideas fostering competitiveness and innovation. Big Data technologies/methodologies have emerged as the solution to this need. However, being a relatively new trend there is still much that remains unknown. This study, based on a risk and benefits perspective, uses the theory of planned behavior to develop a model that predicts the intention to adopt Big Data technologies.

Introduction

Understanding the adoption of information technology (IT) innovations continues to be a challenge for information systems (IS) researchers (Venkatesh, 2006). Every aspect of society, including business and culture, is currently in the midst of a technology-based phenomenon. Advances in digital sensors, communications, mobile networks, storage, processing and cloud computing have given rise to huge collections of data, capturing valuable information to business, science, governments, and society (Bryant et al. 2008, Firestone 2010). By 2020, more than 2.7 zettabytes of data will be created annually reaching 35 zettabytes (IDC 2011) this will call into question the ability of firms to analyze information. Traditional decision-making systems are incapable of adequately resolving this problem. Therefore, companies are starting to roll out their own Big Data initiatives and building massive database systems to drive significant new growth in their business operations (Manyika et al., 2011).

Although the concept of Big Data exists since 2001 when the META Group analyst Doug Laney (Laney 2001) defined data growth challenges and opportunities as being three-dimensional, i.e. increasing *volume* (amount of data), *velocity* (speed of data in and out), and *variety* (range of data types and sources), only in the last two years Big Data has become one of the IT industry's hottest topics. In the press literature, Big Data is characterized as the new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high velocity capture, discovery and/or analysis (Woo et al. 2011).

The Big Data market is expanding rapidly since many firms are expending significant resources on related projects, or are planning to. According to IDC (2012), this market is expected to grow from $3.2 billion in 2010 to $16.9 billion in 2015 based on the premise that these technologies will improve operational efficiency and drive innovation.

Software Vendors such as IBM, Oracle, Microsoft, EMC or SAP, are already providing Big Data services as a source of competitive advantage for their customers.

Big Data systems are being implemented in multiple industries, including commerce, science, and society (Bryant et al. 2008), but many companies still are not interested in this new trend. A Big Data survey conducted in June 2012 by IDC found that 47% of 502 companies across different industries think that they do not need Big Data technologies and 25.8% of them do not see the value it can generate for their companies. Simon (2010) provides a sobering statistic: three out of five Big Data projects do not meet expectations in terms of cost and performance. The major implementation costs are incurred during the integration of Big Data into the existing IT framework. Also, given the high level of sophistication required for Big Data projects (Mckinsey 2011), there are some fears related to the implementation playing against adoption.

All together, these facts lead to the conclusion that the market is at an early stage of adoption, hence only early adopters are betting on these new technologies.

Overall, Big Data represents a disruption in decision-making by enabling business processes to be effectively based on information. Nonetheless, the main challenge at this point is not the deployment of the technology, but rather the transformation of the culture, processes, and people within organizations.

The overall purpose of this study is to explore the impact of Big Data technologies perceived risks and benefits in the intention to adopt them. Since behavioral intention may not be reflected in actual use, this paper also examined the relationship between intended and actual use.

Theoretical background

The academic literature on Big Data is still scarce. Recent articles published focus more on the software, algorithms and hardware needed for Big Data, especially in techniques such as Hadoop, while the adoption decision issues remain unattended.

The initial definition of Big Data was composed of three-dimensional characteristics (known as the 3vs model): volume, variety and velocity. Volume refers to the need for intensive and complex processing of data subsets that actually contain information of value for an organization. Variety refers to the combination of different types of data from different sources. The attribute of variety therefore alludes to the fact that data can come from inside or outside the organization, and may also be structured, semi-structured, or unstructured. Finally velocity, not all of the data in an organization has the same urgency of analysis. There is a full range of velocities: from data that can be batch processed (as in the case of data warehousing) to data that must be processed in real time (when continuous data streams need to be analyzed). The key to understanding speed in Big Data is to clearly identify the informational requirements of the processes and business users.

In 2012, Gartner updated its definition as follows: "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." (Laney 2012).

Perceived benefits of big data

There is a fourth characteristic for Big Data: Value. In the context of Big Data, value refers to: (1) the cost of the technology, which has dropped to allow more companies to undertake this type of projects, and (2) the benefits generated by the use of Big Data (cost reduction, operational efficiency, and business improvements and new revenue streams).

Like any other new technologies, Big Data comes with benefits and drawbacks. Table 1 presents a list of several key benefits and risks developed by McKinsey Global Institute (2011).

Benefits	Risks
Creating transparency by making data accessible to relevant stakeholders in a timely manner Improve operational efficiency (cost, revenue and risk) Use data and experiments to expose variability and raise performance Segment populations to customize the way your systems treat people Use automated algorithms to replace and support human decision making Innovate with new business models, products, and services Sector-specific business value creation	Data quality Talent scarcity (lack of data scientists) Privacy and security concerns Big Data integration capabilities Decision-making Organizational maturity level

Table 1: Perception of benefits for Big data

Decomposed theory of planned behavior

Decomposed Theory of Planned Behavior (DTPB) was raised by Taylor and Todd in 1995. DTPB is an extension of the Theory of Planned Behavior (TPB) developed by Ajzen (1988, 1991). TPB encompasses

three constructs, the attitude toward the behavior, subjective norm, and perception of behavioral control – that when combined form behavioral intention. Intention is then assumed to be the immediate antecedent of behavior (Ajzen 2002). Table 2 presents brief descriptions of the constructs used in TPB.

H14. Perceived behavioral control has a positive effect on actual adoption of Big Data. H15. Intention to adopt Big Data has a positive effect on actual adoption of Big Data.

Antecedents of big data adoption intention

Based on DTPB, in our research model Big Data adoption intention is jointly determined by the individual's Big Data Attitude, subjective norms, and Perceived Behavioural Control. Thus we hypothesize:

Construct	Definition
Behavioral Intention	Refers to individual's intention to perform a behavior and is a function of Attitude, Subjective
Attitude	Refers to individual's positive or negative evaluation of the behavior (Ajzen,
Subjective Norm	Refers to individual's "perception of social pressure to perform or not to perform the behavior" (Ajzen, 1988, p.132)
Perceived Behavioral control	Refers to the "perceived ease or difficulty of performing the behavior and reflects past experience as well as anticipated impediments and obstacles" (Ajzen, 1988, p.132)

Table 2: Definitions of predictors of behavior in the theory of planned behavior (TPB)

Taylor and Todd (1995) also specified that, based on the diffusion of innovation theory, the attitudinal belief has three salient characteristics that influence adoption; relative advantage, complexity and compatibility (Rogers, 1983). Relative advantage refers to the degree to which an innovation provides benefits superseding those of its precursor. This may incorporate factors such as economic benefits, image, enhancement, convenience and satisfaction (Rogers 1983). Complexity represents the degree to which an innovation is perceived to be difficult to understand, learn or operate (Rogers, 1983). The complexity construct is extremely similar, although it is conceived in the opposite direction as ''perceived ease of use'' (Technology acceptance model, Davis 1989). Innovative technologies that are perceived to be easier to use and less complex have a higher possibility of acceptance and use by potential users. Thus, complexity would be expected to have negative relationship to attitude. Complexity (and its corollary, ease of use) has been found to be an important factor in the technology adoption decision (Davis et al. 1989).

Theoretical model and research hypotheses

Synthesizing the theoretical background, we propose the following model (see figure 1) based on DTPB for understanding factors influencing Big Data adoption.

Antecedents of Big Data Adoption

Based on DTPB, the adoption adopt Big Data will be determined by intention to adopt Big Data and perceived behavioral control. As a consequence, we hypothesize:
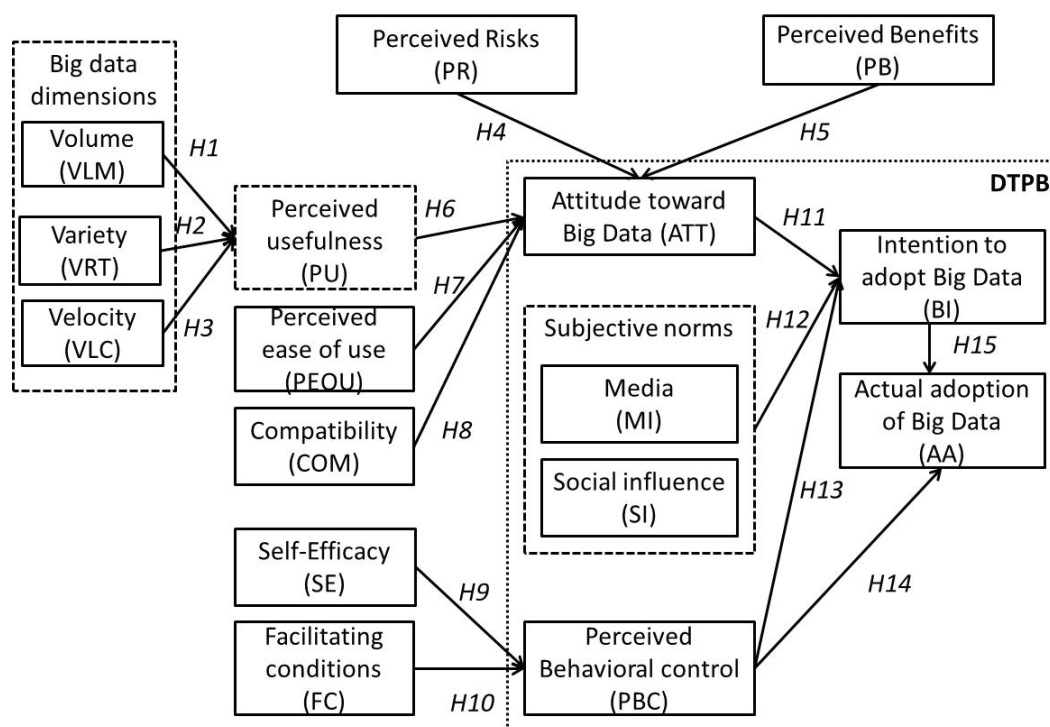
Figure 1: The proposed research model and research hypotheses

H11. Attitude towards Big Data has a positive effect on intention to adopt Big Data. H12. Subjective norm has a positive effect on intention to adopt Big Data.

H12.1. Media has a positive effect on intention to adopt Big Data. H12.2. Social influence has a positive effect on intention to adopt Big Data.

H13. Perceived behavioral control has a positive effect on intention to adopt Big Data.

Antecedents of attitude

Big Data requires of technologies that process and analyze large amounts of heterogeneous data within the right scope of time. These technologies includes A/B testing, association rule learning, classification, cluster analysis, crowdsourcing, data fusion and integration, ensemble learning, genetic algorithms, machine learning, natural language processing, neural networks, pattern recognition, predictive modeling, regression, sentiment analysis, signal processing, supervised and unsupervised learning, simulation, time series analysis and visualization, Massively Parallel-Processing (MPP) databases, search-based applications, data-mining grids, distributed file systems, distributed databases, cloud computing platforms, the Internet, and scalable storage systems. Depending on the degree of knowledge of these technologies, an organization may consider that Big Data is more or less easy to use.

It is reasonable to infer that the perceived ease of use positively influence the company's perceived usefulness and intention to adopt Big Data. Therefore, we hypothesize that:

H7. Perceived ease of use has a positive effect on attitude towards Big Data.

Perceived Usefulness is defined as the degree to which a person believes that adopting Big Data would enhance his or her job performance (Davis 1989). Therefore, we hypothesize that:

H6. Perceived usefulness has a positive effect on attitude towards Big Data

Also, as previously discussed, there are three main reasons to Big Data adoption, namely: volume, variety and velocity. Thus we hypothesize:

H1. Volume has a positive effect on perceived usefulness towards Big Data H2. Variety has a positive effect on perceived usefulness towards Big Data H3. Velocity has a positive effect on perceived usefulness towards Big Data.

As discussed in section 2.1, Big Data generates many potential benefits for companies such as cost control, revenue generation, risk control, decision-making improving, etc. Therefore, it is reasonable to infer that Big Data Technologies perceived benefits positively influence the company's attitude and intention to adopt Big Data.

H5. Perceived benefits have a positive effect on attitude towards Big Data.

Similarly, it is reasonable to infer that the perceived risks of Big Data negatively influence the company's attitude and intention to adopt Big Data. Among them: Talent scarcity, organization maturity, Big Data internal capabilities and data quality.

H4. Perceived risk has a negative effect on attitude towards Big Data.

Compatibility is the degree to which the innovation fits with the potential adopter's existing values, previous experience and current needs (Rogers, 1983). Tornatzky and Klein (1982) found that an innovation is more likely to be adopted when it is compatible with the job responsibilities and value system of the individual. Therefore, it may be expected that compatibility has a positive influence on Big Data adoption. The existence of information systems such as e-commerce platforms, Enterprise Resource Planning (ERP), Business Intelligence (BI), Customer Relationship Management (CRM) or product lifecycle management (PLM), external sources of information and the need to make decision near real-time are factors that generate Big Data situations. It is reasonable to infer that compatibility has a positive influence on attitude towards Big Data. Hence, we hypothesize:

H8. Compatibility has a positive effect on attitude towards Big Data.

Antecedents of perceived behavioral control

According to Ajzen (1988), Perceived Behavioral Control reflects beliefs regarding access to the resources and opportunities needed to perform behavior, or alternatively, to the internal and external factors that may impede performance of the behavior. This notion encompasses the component of "facilitating conditions" (Triandis 1980) and self-efficacy (Bandura 1982). In this research, we define Perceived Behavioral Control as the degree to which external and internal factors influence, knowledge-seeking behavior in an EKR. Thus, we hypothesize:

H9. Self-efficacy has a positive effect on Perceived behavioral control to adopt Big Data.

H10. Facilitating conditions have a positive effect on Perceived behavioral control to adopt Big Data.

Research methodology

Data for this study was collected using an online survey questionnaire. The participants in the survey were managers involved in Big Data adoption decision and usage such as CIOs, marketing directors, and business analytics managers.

Based on the list of the top 100 Spanish companies firms, we contacted the users through email and/or Linkedin. The questionnaire has two parts. The first considers demographic information with control variables such as the job role of the participant, size of the company, and existence of a data mining data center. The second part considers the theoretical model. The measurement items in the questionnaire were developed for the decision variables of attitude, perceived behavioral control, intention to adopt, and actual adoption by adapting the measures proposed and validated by Azjen (2002) to fit the Big Data context. The total number of answers was 53. Table 3 reports the demographic breakdown of the research sample.

Variable	Sub-category	Number (n=53)	%
Business sector	Services	15	28.3
	Public sector	11	20.75
	Manufacturing	2	3.77
	Education	2	3.77
	Health/Pharmaceutical	3	5.66
	Banking/Finance	7	13.21
	Other	13	24.53
Functional Area	Technology	27	50.94
	Marketing/sales	7	13.21
	Operations	4	7.55
	Finance	3	5.66
	Top management	3	5.66
	Other	8	15.09
Annual Revenue	>10 million euros	13	24.53
	10 to 50 million euros	5	9.43
	>50 million euros	35	66.04

Table 3: Research sample demographics

A SEM technique was used to examine the relationships among the constructs. The Partial Least Squares (PLS) approach was chosen for its capability to accommodate small-sized samples (Chin 1998). Further, PLS recognizes two components of a causal model: the measurement and the structural model. Additionally, PLS is especially suitable for exploratory research focusing on explaining variance. Given the aforementioned PLS seemed particularly relevant for this exploratory study – one that is limited by sample size.

Construct reliability and validity

Table 4 shows the factor loadings, Cronbach's alphas (A), Average variance extracted (AVE), and R^2 values. All Cronbach's alphas exceeded the recommended minimum value of 0.7 with the exception of perceived risks variable and, all of the observed construct reliabilities (C.R.) were higher than 0.8 (Fornell and Lacker 1981) with the exception of perceived risks variable. All construct loadings were found to be significant at greater than the recommended p-value of 0.05 (Gefen and Straub 2005) and typically exceeded the recommended

threshold value of 0.707 (Barclay et al. 1995) with the exception of perceived risk, perceived benefits and behavioral intention that were inferior in some constructs. Average variance extracted (AVE) was found to account for a minimum of 50 percent of the variance in each construct and the square root of AVE for each construct was much larger than the construct's correlation with every other construct (Barclay et al. 1995; Gefen and Straub 2005). Measurement items loaded on their respective constructs at a value of at least 0.1 greater than their loading on other constructs (Barclay et al. 1995; Gefen and Straub 2005) and all items loaded higher on their intended construct than on any other construct. Hence, it was concluded that the construct measurement items were consistent and exhibited a substantial degree of convergent and discriminant validity.

Factor	Item	Loadings	AVE	Cronbach	Composite Reliability	R^2
ATT	ATT1	0.972				
	ATT2	0.976	0.946	0.712	0.981	0.407
AA	-	-	1.000	1.000	1.000	0.568
	PB1	0.838				
	PB2	0.524				
	PB3	0.653				
	PB4	0.459				
	PB5	0.609				
	PB6	0.848				
PB	PB7	0.790	0.475	0.812	0.859	-
	BI1	0.975				
BI	BI2 BI3	0.976	0.952	0.95	0.975	0.476
	C1	0.833				
	MI1	0.889				
MI	MI2 MI3	0.896	0.782	0.862	0.915	
	PBC1	0.859				
	PEOU1	0.897				
PEOU	PEOU2	0.936	0.875	0.933	0.954	-
	PU1	0.861				
	PU2	0.926				
	PU3	0.897				
	PU4	0.931				
PU	PU5	0.823	0.789	0.933	0.949	0.283
	PR1	0.632				
	PR2	0.010				
	PR3	0.790				
	PR4	0.122				
PR	PR5	-0.627	0.286	0.30	0.194	-
	SE1	0.827				
SE	SE2 SE3	0.965	0.84	0.904	0.94	-
	SI1	0.954				
SI	SI2 SI3	0.935	0.83	0.896	0.936	-
	FC1	0.912				
	VLC1	0.756				
	VLM1	0.895				
VRT	VRT1	1.000	1.000	1.000	1.000	-

Table 4: Convergent, discriminant validity and reliability of measurements

Path analysis

SmartPLS (Version 2.0.M3) (Ringle et al. 2005) was used to evaluate the statistical significance and relative salience of the research hypotheses. Results of model testing indicated that the constructs included in the research model accounted for approximately 47.6 percent of the variance in the intention to adopt Big Data and 56.8 percent of the variance in actual use of Big Data (Figure 2). Chin (1998) notes that path coefficient values between 0.20 and 0.30 are adequate for meaningful interpretations. Thus, in particular, the results provided support for the significance of eleven research hypotheses. R^2 values, which indicate the predictive power of the model, ranged from 0.28 to 0.7, indicating that the fit of the research model was acceptable.
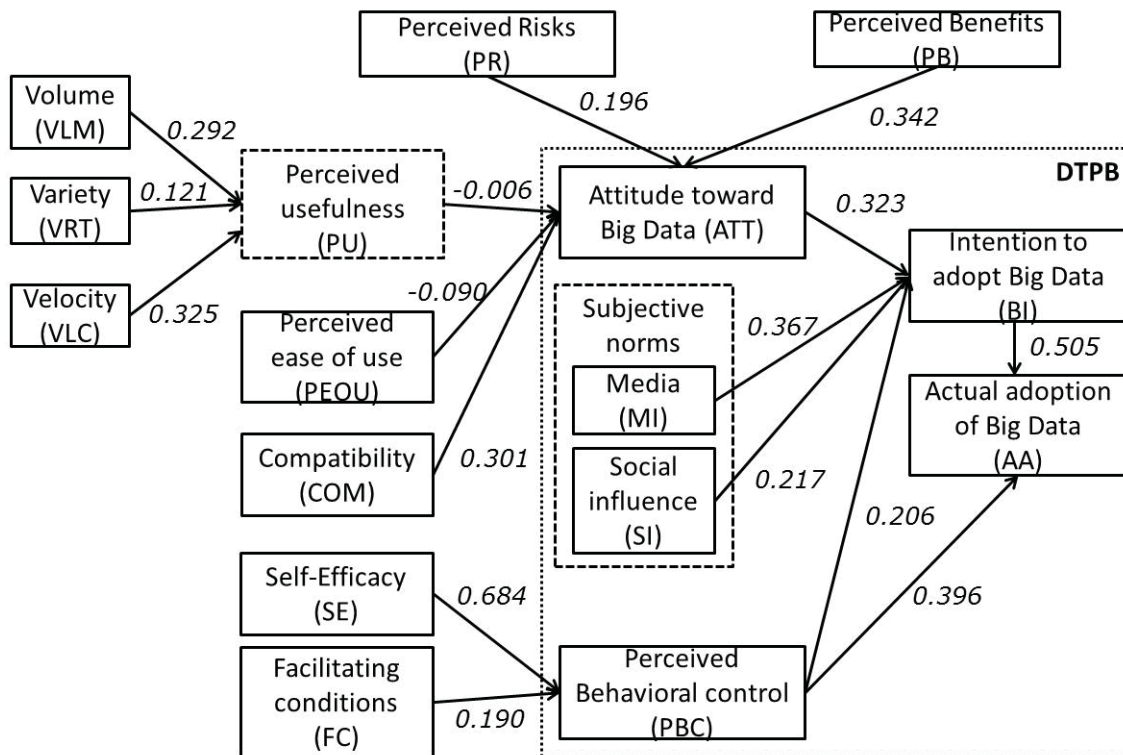
Figure 2: Main study path model results

Discussion

Adding to previous literature on Big Data, the first contribution of this study is the recognition that volume and velocity are the key aspects in Big Data adoption and they have a significant impact in the intention to adopt these technologies. Although, Variety seems not having still such effect, it is expected to become an important factor in determining adoption. The logic behind is that the more heterogeneous and unstructured the data is, the higher the barriers to capture and analyze data. What is clear is as corporate systems are built into Database Management Systems (DMBS), companies perceive volume and velocity as more urgent matters than variety. Also, companies have traditionally focused more on numerical and structured data rather than working with different types of data. However, with the increasingly diversity of data, being able to manage that aspect will play a key part in companies´ data strategy.

Even though the traditional definition of perceived usefulness does not have an impact on the attitude toward Big Data, our model shows that perceived benefits have a significant impact on behavior. Thus, in the subsequent/confirmatory study we

plan to use perceived benefits as the construct that replaces perceived usefulness.

Regarding perceived risks, the exploratory results suggest that perceived risks variable and measurements need to be re-defined. Construct loadings are not statistically relevant, so we need to adjust the constructs definition. Hence, the definition of the potential Big Data risks needs to be reviewed and perhaps extended with more risks. However, the results lead to the belief that perceived risks might have a moderate effect on the attitude towards Big Data adoption. Finally, our results suggest that Media and press news about Big Data have a stronger impact on the decision to adopt Big Data than social influences (friends and/or colleagues suggestion to adopt Big Data).

Therefore the results indicate that specific opportunities as well as challenges exist in Big Data technologies adoption.

Considerations and future work

This research-in-progress contributes to the existing body of knowledge on Big Data by developing a theoretical model to explore and predict the intention to adopt Big Data technology. By extending the theory of planned behavior with the concepts of perceived benefits, risks and

perceived usefulness of Big Data, we seek to understand the adoption of Big Data. Overall, our exploratory results suggest that the proposed model is a first fruitful step to design a theoretical model to predict Big Data adoption.

Also, our exploratory model provides insightful evidence to further research and analysis, especially in terms of perceived risks and the variables that impact on the attitude to adopt Big Data such as velocity and volume.

As a future work, we will review the literature on Big Data risks and redesign the perceived risks construct and then conduct a confirmatory study with a bigger sample size.

References

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. Journal of Applied Social Psychology, 32, pp 665–683.

Ajzen, I. (1991). The theory of planned behavior, Organizational Behavior and Human Decision Processes, vol. 50, pp 179-211.

Ajzen, I. (1988). Attitudes, personality and behavior. Milton Keynes: Open University Press.

Bandura A. (1982). Self-Efficacy Mechanism in Human Agency. American Psychologist, No. 37, pp 122-147.

Bantleman, J. (2012, April 16). The big cost of Big Data. In E. Savitz, CIO network: Insights and ideas for technology leaders [Web log post]. Forbes Magazine. Retrieved October 4, 2012 from http://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/ .

Barclay, D. W., Higgins, C. A., & Thompson, R. (1995). The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adaptation and Use as an Illustration, Technology Studies, 2(2), 285-309.

Benjamin Woo, Dan Vesset, Carl W. Olofson, Steve Conway, Susan Feldman, Jean S. Bozman. (2011). Worldwide Big Data Taxonomy, IDC report.

Bryant, R. E., Katz, R. H., & Lazowska, E. D. (2008). Big-data computing: Creating revolutionary breakthroughs in commerce, science and society. Computing Research Association.

Chin, W. (1998). Issues and opinion on structural equation modeling, MIS Quarterly, 22(1), 7-16.

Dan Vesset, Benjamin Woo, Henry D. Morris, Richard L. Villars, Gard Little, Jean S. Bozman, Lucinda Borovick, Carl W. Olofson, Susan Feldman, Steve Conway, Matthew Eastwood, Natalya Yezhkova. (2012). Worldwide Big Data Technology and Services 2012–2015 Forecast, IDC.

Davis, F.D. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science, No. 35, pp 982-1003.

Deloitte (2012). Billions and billions: Big Data becomes a big deal, Deloitte, http://www.deloitte.com/view/en_GX/global/industries/technology-media-telecommunications/tmt-predictions-2012/technology/index.htm

Firestone, C. (2010). Foreword. In D. Bollier, The promise and peril of Big Data (pp. vii - ix), Washington, DC: The Aspen Institute, https://www.c3e.info/uploaded_docs/aspenbig_data.pdf .

Gefen, D., & Straub, D. (2005). A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example, Communications of the Association for Information Systems, 16(1), 91-109.

Hurwitz, J. (2012, Apr 30). The big deal about Big Data. Business Week, , 1. http://www.businessweek.com/articles/2012-04-23/the-big-deal-about-big-data .

Kiron, D. (2012). All fired up in massachusetts: The states new wave of Big Data companies. MIT Sloan Management Review, Vol. 53, No. 3, pp 1-3.

Lamont, J. (2012). Big Data has big implications for knowledge management. KM World, 21(4), 8-11.

Laney D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Laney D. (2012). Douglas, Laney. The Importance of 'Big Data': A Definition.

Gartner. http://www.gartner.com/resId=2057415

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute.

Mathieson, K. (1991). Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior, Information Systems Research, vol. 2, No. 3, pp 173-191.

Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS 2.0 (M3) Beta, Hamburg, Germany: University of Hamburg (http://www.smartpls.de).

Rogers E. M. (1983). Diffusion of Innovations (3rd edition). London: The Free Press.

Simon, P. (2010). Why new systems fail. Boston, MA: Course Technology, a part of Cengage Learning. Strenger, L. (2008). Coping with Big Data Growing Pains. Business Intelligence Journal, vol. 13, No. 4, pp 45-52.

Taylor, S. & Todd, P. (1995). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. International Journal of Research in Marketing, 12, 137-156.

Tornatzky, L.G., & Klein N. (1982). Innovation characteristics and innovation adoption implementation: A meta-analysis, IEEE Transactions on Engineering Management, 29, pp 28-45.

Triandis, H.C. (1980). Beliefs, Attitudes and Values. University of Nebraska Press, Lincoln, NE.

Venkatesh, V. (2006). Where to go from here? Thoughts on future directions for research on individual-level technology adoption with a focus on decision-making. Decision Sciences, vol. 37, No. 4, pp 497–518.

White, M. (2011). Big Data-Big Challenges. Econtent, vol. 34, No. 9, pp 21.