RESEARCH ARTICLE OPEN ACCESS

Future Prospects of Data Science in Corporate Social Responsibility in India Using in Business Intelligence

Kotadi Chinnaiah[1], Akash Saxena [2]

[1](Rescarch Scholor, Depaptment Cse, Sunrise University, Alwar, Rajasthan, India)

[2](Rescarch Superviser, Depaptment Cse, Sunrise University, Alwar, Rajasthan, India)

Abstract:

Agile methodologies were introduced in 2001. Since this time, practitioners have applied Agile method-ologies to many delivery disciplines. This article explores the application of Agile methodologies and principles to business intelligence delivery and how Agile has changed with the evolution of business intelligence. Business intelligence has evolved because the amount of data generated through the internet and smart devices has grown exponentially altering how organizations and individuals use information. The practice of business intelligence delivery with an Agile methodology has matured; however, business intelligence has evolved altering the use of Agile principles and practices. The Big Data phenomenon, the volume, variety, and velocity of data, has impacted business intelligence and the use of information. New trends such as fast analytics and data science have emerged as part of business intelligence. This paper addresses how Agile principles and practices have evolved with business intelligence, as well as its challenges and future directions.

1. Introduction

The manifesto and principles for Agile Software Development (ASD) were published in 2001, and since then, the objectives and principles have been interpreted and applied to Business Intelli-gence (BI). The application to BI is natural, because of the iterative and incremental nature of BI development. The intent of this article is to provide practitioners an understanding of how the Agile prin-ciples can be applied to BI delivery, fast analytics, and data science. Beck et al. (2001) outlined the core ideals of the manifesto: individ-uals and interactions over processes and tools; working software over comprehensive documentation; customer collaboration over contract negotiation; and responding to change over following a plan. The result of following these ideals, software development becomes less formal, more dynamic, and customer focused.

Information Technology (IT) departments are faced with main-taining a competitive edge, which, in turn, increases pressure to deliver high quality technology solutions faster. Under these cir-cumstances, the value of technology efforts are determined based on how soon payback and return on investment occur. BI initiatives require significant upfront and ongoing investment to maintain value, inviting constant scrutiny on whether business value occurs. Measuring BI value continues to be a struggle for organizations, mainly due to the challenge of directly attributing return to the investment in BI. BI plays the role of an enabler – enabling the organization to become smarter, work smarter, and make better decisions through the use of information. The enabler role makes it difficult to directly attribute a return on investment and over time, the use of information becomes routine and expected.

The information value chain is the process used to derive value from information and information from data; BI delivery is cen-tered on the information value chain. Collecting raw data is the first step in the value chain; applying logic and business context to the data creates information; information is then consumed by BI users; decisions and actions are a result of the consumption of data; resulting in decisions and actions that provide business value.

The information value chain is an important concept in under-standing the benefits of Agile principles applied to BI delivery. BI delivery is not accomplished via traditional waterfall software development (although some organizations attempt this); it is more focused on data discovery and understanding how informa-tion is going to be used. This perspective drives how Agile principles should be applied to BI delivery – less focus on software devel-opment and more focus on information use. The need to delivery faster has increased over the last 5 years due to the demand of real-time data analysis (Halpern, 2015). The Internet of Things (IoT), where data collection is embedded into devices, contributes to this demand for fresher data. Monitoring equipment failures, for exam-ple, will be possible with data that is seconds old versus data that is hours or days old (Halpern, 2015).

The objectives of this article are fourfold. First, revisit the align-ment between Agile principles and BI delivery, fast analytics, and data science. Second, analyze Agile methodologies and how they have been applied with BI and are emerging with Big Data. Third, review the components and best practices of Agile BI delivery con-sidering the impact of Big Data. Last, propose an Agile framework for BI delivery, fast analytics, and data science; fast analytics and data science are the emerging data analysis trends due to Big Data (Fig. 1).

2. RELATEDWORK

Business Intelligence (BI) is defined by literature and scholars in similar ways. Noble (2006) defines BI as the ability to provide the business an information advantage; business doing what it has always done, but more efficient. Singer (2001) described BI as the value proposition that helps organizations tap into decision-making information that regular reporting does not provide. Singer outlined that BI requires tools, applications, and technologies focused on enhanced decision-making and is commonly used in supply chain, sales, finance, and marketing. Negash and Gray (2008) outlined BI more comprehensively. BI is a data driven process that combines data storage and gathering with knowledge manage-ment to provide input into the business decision making process. BI enables organizations to enhance the decision making process and requires processes, skills, technology, and data. More recently, Gartner (2013) and Halpern (2015) have extended BI to be an umbrella term which includes applications, tools, infrastructure, and practices to enable access and analysis of information to opti-mize performance and decision-making.

The challenges In BI delivery include business and IT collabo-ration that results in data become information. Delivery of BI is accomplished via a methodology. Creswell (2005) outlined that a methodology is set of processes, methods, and rules applied within a discipline. Successful BI methodology should focus on the infor-mation value chain and less on the development of software as is the focus of traditional information technology (IT) development. Research has demonstrated that waterfall lifecycles and traditional software development practices are not successful in BI. Software and hardware do not provide organizations value pertaining to BI; it is the use of information (Larson, 2009).

Common stumbling blocks traditionally experienced in BI projects included: fuzzy requirements; lacking an understanding about how data is created and used; data quality is not measured or known; source system constraints dictate design and service levels; developing based on perceptions of data; results are not demonstrated in a timely manner; and working with a lack of trust between IT and business stakeholders (TDWI, 2009). While these challenges still remain, the need to have information sooner has been influenced by the phenomenon of “Big Data”. Big Data is a broad term used to describe data sets that are large, complex, and cannot be addressed by traditional IT methodologies and applica-tions (Davenport, 2013).

2.1. Big data

Traditional data processing has transformed because how data is generated today has transformed (Davenport, 2013). Historically IT departments managed transaction processing systems. Business was about transactions – orders, sales, shipments, inventory, and accounting, to list a few examples. Transactional data is structured,stable, and understood by organizations. Structured data is format-ted in rows and columns. Transactional data is primarily used for decision support. The differentiating points between transactional data and Big Data are volume, variety, and velocity. Volume refers to the amount of data, variety is based on the types of data sources, and the velocity represents the age of data. Volume, variety, and velocity are referred to as the "3 V's" (Davenport, 2014). Trans-actional data is structured, well understood, and volume is tens of terabytes or less (Davenport, 2014). Transactional data is used in decision support analysis. Decision support analysis focuses on "what has happened" or retrospective analysis.

Volume for Big Data is measured in excess of 100 TB or petabytes (although research in this threshold varies). Big Data is character-ized by data types considered unstructured – not predefined or known – thus a high degree of variety. Velocity of Big Data can be measured in less than a few minutes. Big Data is used in machine learning and predictive analysis where organizations focus on "what will happen" versus "what has happened" (Davenport, 2014). The emergence of Big Data has changed the face of BI.

2.2. Analytics

Analytics have been around since the 1950s and are not a new idea (Davenport, 2014). Analytics started with a limited number of data sources which came from internal systems and data was stored in a repository such as a data mart or data warehouse—defined as traditional BI. Most analysis was descriptive and BI consisted pri-marily of reporting. In 2003, Big Data started emerging when high technology firms such as Google and Yahoo began using Big Data for internal analytics and customer-focused processes. The veloc-ity of Big Data changed traditional BI as data had to be stored and processed quickly. Predictive and prescriptive analytics started to emerge, but visual analytics of descriptive data was still the promi-nent form of analytics (Davenport, 2014). Big Data became "bigger"– more volume, variety, and velocity – and organizations began focusing on the data-driven economy. Organizations of all types are developing data-based offerings to remain competitive (Davenport, 2014).

With the advent of Big Data and analytics evolving, BI delivery has been impacted. Data has to be turned into information quickly for analysis. Organizations are focusing more on prescriptive and predictive analysis that utilize machine learning as well as fast ana-lytics through visualization. Fast analytics refers to the ability to acquire and visualize data quickly (Halpern, 2015; Jarr, 2015). The velocity increase in data has accelerated the need for IT depart-ments to acquire data and transform it into information. Table 1 illustrates the characteristic differences between traditional BI and fast analytics with Big Data.

3. Application of agile to BI and Big Data

Agile ideals and principles were published by Beck et al. (2001) and since this time, practitioners have focused on applying an Agile approach to BI. The challenges that BI projects face make the Agile approach an attractive answer due to the parallels that exist between them. By using an Agile approach, means the methodology is less formal, more dynamic, and customer focused. The dynam-ics required in BI delivery make Agile approach a good fit with BI; however, practice with Agile methodologies have resulted in adjustments to the well-known Agile methodologies to focus on the utility of information versus primarily software development. Agile methodologies align well also with Big Data where little time is spent defining requirements up front and the emphasis is on devel-oping small projects quickly. Agile methodologies will align well with iterative discovery and validation which support prescriptive and predictive analytics (Ambler & Lines, 2016).

3.1. Agile principles

Analyzing the Agile principles provides an understanding of how using an Agile approach matches well with BI delivery. To reit-erate the principles: individuals and interactions over processes and tools; working software over comprehensive documentation; customer collaboration over contract negotiation; and responding to change over following a plan. Beck et al. (2001) outlined that an Agile approach focuses more on the left side of the principle; however, the right side is not ignored.

3.2. Working software over comprehensive documentation

Documentation is valuable; however, the value is not the issue. Documentation has an inherent problem – usability. Doc-umentation has been a dreaded aspect of traditional development methodologies. Documentation takes too much time to complete, tends to be out-of-date, and is rarely used after the initial deploy-ment. Creating comprehensive documentation does not allow for quick delivery; however, not producing documentation can be more detrimental. For Agile, documentation needs be usable and add value. Documentation should less textual and more visible. Development artifacts in BI such as source to target mappings, dia-grams, and models, are examples of valuable artifacts that are easy to use and maintain. Diagrams can provide a level of documentation that is adequate to support requirements, design, and development and are easy to maintain. A picture is worth a thousand words (Larson, 2009).

Documentation with faster analytics and Big Data aligns with Agile principles and comprehensive documentation is not the norm. Documentation is a lower priority and the priority with faster analytics is insight from information. Due to the speed of data avail-ability, documentation is primarily ignored (Adamson, 2015).

3.3. Customer collaboration over contract negotiation

Practicing ongoing collaboration throughout any process adds value – communication is increased, expectations are consis-tently reaffirmed, and ownership of the end product is shared. Collaboration is emphasized in "interaction and individuals over process and tools" and fundamental to the success of Agile. Without pre-determined expectations, contracts can frame expec-tations but allow refinement and change. The details surrounding requirements are not often known in enough detail to document. Collaboration between stakeholders addresses this via delivery by determining what the expectations are and increasing communi-cation between stakeholders (Larson, 2009).

3.4. Agile methodologies

The manifesto and principles for Agile Software Development (ASD) were published in 2001, and since then, the objectives and principles have been interpreted and applied to new Agile method-ologies. The popular approaches from which the manifesto and principles were derived – Extreme Programming (XP) and Scrum– are in practice today with success and are considered standard development methodologies (Hsieh & Chen, 2015). Agile principles have been applied to other disciplines such as project manage-ment with success (Kaleshovska, Josimovski, Pulevska-Ivanovska, Postolov, & Janevski, 2015). Success with Agile methodologies include reduced cycle time, higher quality, increased require-ments clarify, increased flexibility, and a higher overall stakeholder satisfaction rate when compared to similar projects using dif-ferent project or software development methodologies (Hsieh & Chen, 2015; Kaleshovska et al., 2015). The core practices of Agile methodologies include: small, short releases; stakeholders phys-ically located together; and a time-boxed project cycle (typically 60–90 days, although the cycle may be shorter depending on the deliverable) (Kendall & Kendall, 2004). These practices continue to contribute to the success of Agile projects (Hsieh & Chen, 2015; Kaleshovska et al., 2015).

3.7. Agile and business intelligence

The primary goal of a BI project is to enable the use of informa-tion. If the primary goal of BI is enabling the use of information, then scope of the BI project focuses on turning data into information. Software development is part of the data to information process; however, software development in BI is less about creating a work-ing program and more about application of business context to data. Software used in BI includes database management systems, data cleansing, data transformation, and analytical systems. The scope of development in BI includes more configuration and application of logic versus programmatic coding. In order to understand how to apply logic and configure the software, IT will need to comprehend the business use of data (Larson, 2009). Big Data technology has includes the scope of software and hardware used in BI (Davenport, 2014).

3.8. Agile and Big Data

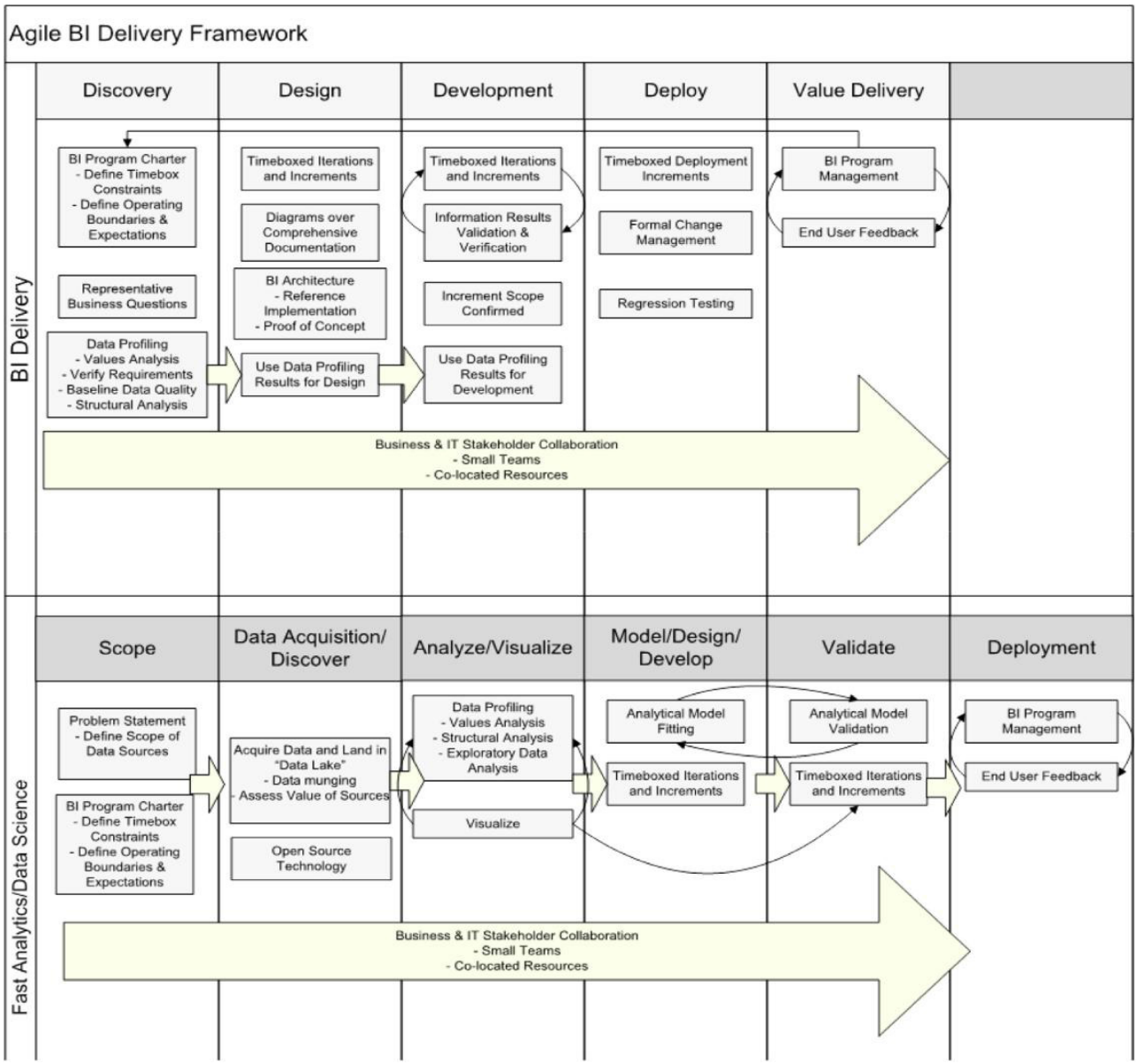
Big Data is a phrase coined to describe the changing technology landscape that resulted in large volumes of data, continuous flow of data, multiple data sources, and multiple data formats (Davenport, 2014). The input to a BI project is data and the output is information. With the data landscape changing so rapidly, BI projects and the methodologies used are also changing.

Big Data is a fairly new phenomenon, thus research is limited. Some research suggest Big Data is a media term or vendor term used to describe hardware, software, and analysis connected with non-traditional data sources (Davenport, 2014). Several analysis trends have emerged (in some cases re-emerged) such as predic-tive and prescriptive analysis. Data science, which is defined as an interdisciplinary field focusing on extracting insights from data in various forms, is the next generation of data analysis fields such as statistics and data mining. With Data Science, analysis is com-pleted using statistics and machine learning algorithms to produce data products or models that perform descriptive, predictive, or prescriptive analysis (Schutt & O'Neil, 2013). As the 3 V's describ-ing Big Data suggest, data freshness, data variety, and the speed of analysis have changed how BI projects have traditionally been approached.

4. Business intelligence delivery, analytics, and Big Data

4.1. Goals of BI delivery

Yeoh and Koronios (2010) posited that a BI system is not a con-ventional IT system (i.e. transactional system); however, BI systems have similar characteristics to enterprise systems or infrastructure projects. BI system implementation is a complex activity involving hardware, software, and resources over the life of the system. The complexity of the BI system infrastructure increases with the scope. An enterprise BI system can include a data warehouse, integrated data structures, source systems, and large data volumes.



4.2. Iteration and incremental

One of the synergies that Agile has with BI is the short, small release and experts' recommendation that BI, fast analyt-ics, and data science projects work best delivered in increments (Davenport, 2014; Yoeh & Koronios, 2010; Mungree, Rudra, & Morien (2013)). This incremental approach supports that fact that businesses and technology change quickly and want to evaluate the impact of these changes. An incremental approach allows for management of risk, allows for more control, and enables customers to see tangible results.

Table 2

Comparison of the business intelligence lifecycle and fast analytics/data science project lifecycle.

Business Intelligence Lifecycle	Fast Analytics/Data Science Lifecycle
Discovery	Scope
Design	Data Acquisition/Discovery
Development	Analyze/Visualize
Test	Model/Design/Development
Deploy	Validate
Support	Deployment
Value	Support/Feedback

Increments deal with the staging and scheduling of deliverables which may occur at different rates. Iterations are cycles to revise and improve the deliverable. Increments are scheduled as part of a roadmap or release plan tied to an overall BI strategy that outlines what information capabilities are needed and when. Iterations will happen within the increment. Increments are time-boxed, there-fore the results can be less or more than expected. If less than expected is delivered, increments are adjusted accordingly. Sim-ply, increments manage the scope of the delivery and iterations are used to refine the quality of the deliverable. Deliverables can be code, models, diagrams, or any artifact created as part of the cycle.

4.3. The BI and fast analytics lifecycle

A lifecycle is the progression of something from conception to end of life or when something no longer provides value. Lifecy-cles have phases that comprise the progression of conception to end; the BI lifecycle is no different. The BI lifecycle parallels the SDLC with similar phases and is centered on the utility of informa-tion versus the development of software. Fast analytics and data science projects take on a different approach due to the speed of technology and the acquisition of data. In BI, discovery focuses on requirements, design on defining and working with structured data, development on creating code and databases, testing on vali-dating development, deployment to production, then support and value measurement. With fast analytics and data science, design is encapsulated into development, because data is acquired too quickly to analyze and it is unstructured and dynamic. Fast ana-lytics can involve iteration and visualization of data to understand and define. Data science involves iterative development of analyt-ical models where models are created, validated, and altered until the desired results are achieved (Schutt & O'Neil, 2013). Table 2 compares the two different lifecycles.

4.4. Data Acquisition/Discovery

New technologies have made it possible to acquire data without a full understanding of its structure or meaning which is the oppo-site of what occurs in the BI lifecycle where data is profiled and analyzed to understand its meaning before loaded into a data repos-itory for use. Hadoop or the Hadoop File System (HDFS) originated at Google and is now used in an open-source format by organiza-tions to land data without the need for data modeling (Davenport, 2014). Analysts use fast analytics to access, assess, and visualize to discover the value and

use of data sources. New data repos-itories such as the "Data Lake" have emerged where technology enables storage and processing power to support analyzing large unstructured data sets (Davenport, 2014).

4.4.1. Analyze/Visualize

For both fast analytics and data science, analysis and visualiza-tion are an iterative process. With fast analytics the primary goal is visual analytics to support analysis. Fast analytics can produce new knowledge that creates a refinement of the visual product. Fast analytic can iteratively produce new dashboards or scorecards to be used in ongoing BI or produce one-time analysis tools to sup-port new knowledge gain. With data science, fast analytics and visualization is completed as part of the exploratory data analysis phase where descriptive analysis is used to highlight variable rela-tionships and identify parameters to be used in analytical models (Schutt & O'Neil, 2013). If fast analytics and visualization produces a BI product such a dashboard or scorecard, the BI product is then validated. It is possible that fast analytics is primarily focused on discover, and a BI product is not produced.

4.4.2. Agile delivery framework for BI, fast analytics and data science

This author compiled a BI Delivery Framework based on research and experience, which synthesized Agile practices with BI delivery practices (Larson, 2012). Since the development of this framework, additional research has been published that sup-ports the value of using Agile methodologies with BI projects. Fast analytics and data science has become prominent practices in data analysis due to the emergence of Big Data. Organizations are expanded competencies in the BI field to include data scientists (Davenport, 2014; Schutt & O'Neil, 2013;Mohanty et al., 2013)

Conclusion

This paper focused on the recent developments in adoption of Agile principles to BI delivery and how Agile has changed with the face of BI. Fast analytics and data science have been included under the umbrella of BI. Agile ideals fit well into the BI world and research on successful application has emerged. Agile addresses many of the common problems found in BI projects by promoting interac-tion and collaboration between stakeholders. Close collaboration between parties ensures clearer requirements, an understand-ing of data, joint accountability, and higher quality results. Less time is spent attempting to determine information requirements, and more time is devoted to discovering what is possible. Future research opportunities are abundant as the landscape of BI and data analysis is transforming with Big Data. Topics in discussion have addressed the current challenges and future directions for adopt-ing business intelligence platforms, applications and services for all types of organizations.

References

1. Adamson, C. (2015). The age of big data modeling. www.tdwi.org
2. Ambler, S. (2006). Agile database techniques: effective strategies for the agile developer. John Wiley & Sons.
3. Ambler, S. W., & Lines, M. (2016). The Disciplined Agile Process Decision Framework.
4. Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., et al. (2001). Manifesto for agile software development.. Available on: http:// agilemanifesto.org/
5. Chang, V. (2014). The business intelligence as a service in the cloud. Future Generation Computer Systems, 37(July), 512–514.
6. Chang, V. (2015). Towards a big data system disaster recovery in a private cloud.Ad Hoc Networks, 35, 65–82.

7. Chang, V. (Chang, V., & Ramachandran, M. (2016). Towards achieving data security with the cloud computing adoption framework. IEEE Transactions on Services Computing, 9(1), 138–151.

8. Chang, V., & Wills, G. (2016). A model to compare cloud and non-Cloud storage of big data. Future Generation Computer Systems, 57, 56–76.

9. Cockburn, A. (2008). Using both incremental and iterative development. The Journal of Defense Software Engineering.

10. Couture, N. (2013). Best practices for adopting agile methodologies for data warehouse development. Business Intelligence Journal, 18(2), 8–17.

11. Creswell, J. W. (2005). Educational research: planning, conducting, and evaluating quantitative and qualitative research (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

12. Davenport, T. H. (2013). Analytics 3.0. Harvard Business Review, 91(12), 64–72. Davenport, T. H. (2014). Big data @ work: dispelling the myths, uncovering theopportunities. Harvard Business Review Press.

13. Gartner IT Glossary. (2013). Definition of business intelligence.. Retrieved from: http://www.gartner.com/it-glossary/business-intelligence-bi/ (accessed on 15.04.16.)

14. Gartner Research. (2015). Gartner says business intelligence and analytics leaders must focus on mindsets and culture to kick start advanced analytics. Gartner business intelligence & analytics summit 2015.

15. Halpern, F. (2015). Next-generation analytics and platforms for business success:tDWI research report. Available on: www.tdwi.org

16. Hsieh, C. Y., & Chen, C. T. (2015). Patterns for continuous integration builds in cross-platform agile software development. Journal of Information Science & Engineering, 31(3), 897–924.

17. Hughes, R. (2013). Agile data warehousing project management: business intelligence systems using Scrum and XP. Waltham, MA: Morgan Kaufmann.

18. Jarr, S. (2015). Fast data and the new enterprise data architecture. O'Reilly Publishing.

Kaleshovska, N., Josimovski, S., Pulevska-Ivanovska, L., Postolov, K., & Janevski, Z.(2015). The contribution of scrum in managing successful software development projects. Economic Development/Ekonomiski Razvoj, 17(1/2), 175–194.

19. Kendall, J. E., & Kendall, K. E. (2004). Agile methodologies and the lone systems analyst: when individual creativity and organizational goals collide in the global IT environment. Journal of Individual Employment Rights, 11(4), 333–347.

20. Kiem, D., Kohlhammer, J., & Ellis, G. (2010). Mastering the information age: solving problems with visual analytics. Eurographics Association.

21. Larson, D. (2009). BI principles for agile development: keeping focused. Business Intelligence Journal, 14(4), 36–41. Retrieved from Business Source Complete database.

22. Mohanty, S., Jagadeesh, M., & Srivatsa, H. (2013). Big data imperatives: enterprise big data warehouse. BI Implementations and Analytics.

23. Muntean, M., & Surcel, T. (2013). Agile BI—the future of BI. Informatica Economica, 17(3), 114–124. http://dx.doi.org/10.12948/issn14531305/17.3.2013.10

24. Mungree, D., Rudra, A., & Morien, D. (2013). Proceedings of the Nineteenth Americas Conference on Information Systems.

25. Negash, S., & Gray, P. (2008). Business intelligence. In F. Burstein, & C. W. Holsapple (Eds.), Handbook on decision support systems 2 (pp. 175–193). Berlin, Heidelberg: Springer.

26. Newman, R., Chang, V., Walters, R. J., & Wills, G. B. (2016). Model and experimental development for business data