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 Intelligence (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 principles can be applied to BI delivery, fast analytics, and data science. Beck et al. (2001) outlined the core ideals of the manifesto: 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. The result of following these ideals, software development becomes less formal, more dynamic, and customer focused.

Information Technology (IT) departments are faced with maintaining a competitive edge, which, in turn, increases pressure to deliver high quality technology solutions faster. Under these circumstances, 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 centered 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 understanding 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 information is going to be used. This perspective drives how Agile principles should be applied to BI delivery — less focus on software development and more focus on information use. The need to deliver 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 example, 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 alignment 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 considering 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. RELATED WORK

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 efficiently. 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 management 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 optimize performance and decision-making.

The challenges in BI delivery include business and IT collaboration 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 information 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 applications (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 formatted 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). Transactional 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 characterized 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 primarily 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 velocity 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 prominent 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 analytics 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 departments 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 dynamics 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 developing 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 reiterate 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. Documentation 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 deployment. 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, diagrams, 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 availability, 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 consistently 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 predetermined expectations, contracts can frame expectations 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 communication 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 methodologies. 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 management with success (Kaleshovska, Josimovski, Pulevska-Ivanovska, Postolov, & Janevski, 2015). Success with Agile methodologies include reduced cycle time, higher quality, increased requirements clarity, increased flexibility, and a higher overall stakeholder satisfaction rate when compared to similar projects using different project or software development methodologies (Hsieh & Chen, 2015; Kaleshovska et al., 2015). The core practices of Agile methodologies include: small, short releases; stakeholders physically 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 information. 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 working 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 predictive 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 completed 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 describing 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 conventional 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 analytics, 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.
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. Lifecycles 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 information 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 validating 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 analytics can involve iteration and visualization of data to understand and define. Data science involves iterative development of analytical 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 opposite of what occurs in the BI lifecycle where data is profiled and analyzed to understand its meaning before loaded into a data repository for use. Hadoop or the Hadoop File System (HDFS) originated at Google and is now used in an open-source format by organizations 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 repositories 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 visualization 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 analytics can iteratively produce new dashboards or scorecards to be used in ongoing BI or produce one-time analysis tools to support 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 relationships 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 supports 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 interaction and collaboration between stakeholders. Close collaboration between parties ensures clearer requirements, an understanding 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 adopting business intelligence platforms, applications and services for all types of organizations.

References