

DATA WAREHOUSE AND BUSINESS INTELLIGENCE IMPLEMENTATION: AGILE AND ITERATIVE THEMES

Troy J. Mueller, Robert Morris University, tjmst41@mail.rmu.edu

ABSTRACT

The growth and success of agile software development across industry and government has led to the application of these methods to other domains. This paper discusses common agile and iterative themes used on data warehouse and business intelligence projects by reviewing and analyzing academic and industry literature, and categorizing key words and phrases associated with agile across common themes.

Keywords: Data Warehouse (DW), Business Intelligence (BI), Agile Data Warehouse, Agile, and Implementation

INTRODUCTION

The large volumes of data generated in the daily operations of private industry and government organizations provide an opportunity to better understand and improve effectiveness and efficiency. To accomplish this, organizations must implement data warehouse and business intelligence (DW/BI) capabilities. For the U.S. Federal government this requires smart investment decisions in an austere budget environment. Once a decision to implement has been made, leaders must decide how to implement. Options include traditional sequential waterfall-like, agile/iterative-like, or hybrid approach. The purpose of this paper is to discuss common agile and iterative themes used on DW/BI projects by reviewing and analyzing academic and industry literature.

Beyond this introduction, this paper contains five sections: definition of key terms, agile software development, data warehouse and business intelligence implementation, agile themes, and a conclusion. The definition of key terms operationalizes selected terms to ensure clear understanding of the context of the paper. The two following sections provide back ground information and the agile themes section provides an illustration of the mapping of agile attributes on DW/BI projects. Finally, the paper closes with a brief conclusion.

Definition Of Key Terms

To ensure clarity and common understanding, a number of terms critical to the context of the topic are defined for the reader. A data warehouse is “a physical repository where relational data are specifically organized to provide enterprise-wide, cleansed data in a standardized format” [15, p. 276]. Business intelligence is “a conceptual framework for decision support. It combines architecture, databases (or data warehouses), analytical tools, and applications” [15, p. 275]. Agile software development is “a software development methodology designed for small to medium size projects with uncertain requirements, which emphasizes people rather than process” [9, p. 445]. The waterfall model is “a software lifecycle model in which all development processes occur sequentially” [9, p. 446]. Agile data warehousing is defined as “a set of principles, behaviors, and techniques that provides our customers a far shorter time to value for quality business intelligence through fast project starts, high business involvement, frequent increments of value, and regularly update planning information” [7, p. 327].

Agile Software Development

The waterfall model surfaced in 1970 as a structured alternative to the more common code-and-fix method [4]. Various interpretations of the model exist containing differences in the number and name of the phases. The particular variation of the model used in this paper is that depicted in Quality Software Project Management by Futrell, Shafer, and Shafer. This waterfall model consists of a sequential set of activities separated into nine phases with the focus on early requirements and design work, and large blocks of documentation in support of this early effort. The phases consist of concept exploration, system exploration, requirements, design, implementation, installation, operations and support, maintenance, and retirement. The sequential nature of the model allows for feedback between phases providing the opportunity to “resolve unexpected problems before proceeding” [9, p. 368].

Agile can be described as a mindset rather than a method or model. The mindset can be better understood when comparing agile to waterfall. The ritual of large volumes of documentation, the illusion of defined upfront requirements, and structured review gate events buries organizations and project teams in process. Agile is “an

attitude, a style, a philosophy of approaching problems and their solutions” [8, p. 11]. The focus is always on the user and the capability required making the user successful. These attitudes and beliefs are defined by the principles espoused in the Manifesto.

Like waterfall, various interpretations and approaches of agile development are available for consideration. For illustrative purposes, the method discussed in this article is the scrum process defined by Schwaber (1995) in his paper presented at the 1995 Object-Oriented Programming, Systems, Languages & Applications (OOPSLA) conference. Schwaber (1995) describes scrum as “an enhancement of the iterative and incremental approach to delivering object-oriented software” [11, p.3]. Scrum consists of three phases with a number of sub-phases. These phases are pregame, game, and postgame [11]. The pregame sub-phases consist of planning and architecture [11]. Planning includes definition of release content, cost, and schedule, and architecture involves design detail including changes to the system architecture [11]. In the game phase, one or more development sprints (sub-phase) are conducted to create functioning software [11]. Schwaber defines a sprint as “a set of development activities conducted over a pre-defined period, usually one to four weeks” [11, p.14]. In the postgame sub-phase of closure, the software is tested, documentation is completed, and the release is prepared for delivery to the customer [11]. The theme of agile development is an intense focus on customer interaction and efficiency.

The need for an alternative to waterfall in software development is well documented. The Standish Group’s 1994 Chaos Report described an industry struggling with high failure rates and extreme cost overruns in the rare instance of arguable success [12]. In 1994, 16% of projects were successful and 53% were failures with 31% being classified as challenged [12]. The 2006 report demonstrated improvement with success at 35%, failure at 46%, and challenged at 19% [13]. It should be noted that the 2006 report is after the release of the Agile Manifesto and the spread of agile methods across the industry. The most recent report illustrates a negative shift with successful projects at 32% and failures at 24%; however challenged projects continued the same trajectory with 44% reported [14]. Unfortunately, the continued reduction in challenged projects is at the cost of reduced success and increased failure. Reviewing this data indicates that the move to agile is not complete and not always smooth.

Data Warehouse and Business Intelligence Implementation

DW/BI implementations have been historically classified in one of two categories, Kimbal or Inmon. Bill Inmon, considered the “father of data warehousing,” advocates for a “top-down development, employing established database development methodologies and tools” [15]. Ralph Kimball proposes the alternative of a bottoms up approach that involves building data marts, one at a time [15]. In software engineering parlance, Inmon’s approach can be viewed as waterfall-like and Kimball’s as agile/iterative-like.

Similar to the forces of change in software development, the need for change regarding DW/BI projects is economical and customer driven. Years of countless studies, such as the Standish Group Chaos Reports, provide empirical evidence to support the move to agile software development. While this level of data collection for DW/BI projects is in its infancy, a 2004 study revealed that only 35% of projects valued at \$12.89 million were declared successful [3].

Hughes (2013) defines agile data warehousing as “a set of principles, behaviors, and techniques that provides our customers a far shorter time to value for quality business intelligence through fast project starts, high business involvement, frequent increments of value, and regularly update planning information” [7, p. 327]. The movement towards a more agile approach for DW/BI is mimicking the early days of the agile software development movement to include development of a manifesto. Hughes hosts a website [6] to augment his books on the topic which includes the Agile Data Warehousing Manifesto.

Analogous to the Manifesto for Agile Software Development, the data warehouse version describes activities and actions that are valued more than others. The Manifesto for Agile Software Development was released by the Agile Alliance on February 13, 2001 and states:

Manifesto for Agile Software Development

1. “We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:
2. Individuals and interactions **over** processes and tools
Working software **over** comprehensive documentation
Customer collaboration **over** contract negotiation
Responding to change **over** following a plan
3. That is, while there is value in the items on the right, we value the items on the left more” [1]

On October 10, 2012, Hughes released the Agile Data Warehousing Manifesto which states:

Agile Data Warehousing Manifesto

1. “Using Agile techniques, we are finding ways to break through the high-cost, slow-delivery, and excessive risk that used to make data warehousing/business intelligence projects frequently fail.
2. Through this work we have come to value the following notions in addition to the tenets of the Agile manifesto:
3. Prompt, sponsor-valued results **over** technical perfection
4. Managing risk **over** eliminating uncertainty
5. Evolving data **over** iterating on application code
6. Appropriate technology **over** an infrastructural monoculture
7. While there is value in the items on the right, the goals listed on the right generally lead to far better results.” [6]

Agile Themes

This study reviewed two academic articles and one published book describing three approaches to DW/BI implementation projects. These publications were reviewed with the goal of identifying themes that aligned with agile approaches, practices or processes. Five categories of themes (see Table 1 – Common Themes) emerged; these include iterative requirements, modularity/decomposition of projects, time boxing of capability delivery, team structure, and user involvement. Key words and phrases common to agile were categorized across the common themes.

Golfarelli, Rizzi and Turrinchia explore the benefits of utilizing modern software engineering practices on data warehouse projects and offer a potential design methodology that leverages agile methods [5]. Problems encountered on data warehouse projects include (1) “uncertain & changing requirements,” (2) “linear approach to design,” and (3) “design complexity” [5, p. 67]. Each of these challenges exists on traditional software projects and, as previously addressed in this paper, has inspired change. Four-Wheel-Drive (4WD) is the solution put forward by the researchers, and applies a balanced approach of agile principles and methods [5].

In the article *Developing Realtime Business Intelligence Systems The Agile Way*, Dasgupta and Vankayala use the corporate Global Delivery Model developed by Infosys Technologies to demonstrate the benefits of using agile-like methods on business intelligence implementation projects [2]. The Infosys project concerned “developing and sustaining a complex business intelligence system” which is packaged and sold to the public [2, p. 5]. An added layer of complexity on this project was the geographical separation of the two teams across time zones which were mitigated by overlapping scheduled work hours [2].

The third publication reviewed for this study does not address an actual project; rather it provides guidance that data warehouse implementation teams can utilize to take advantage of the benefits of agile methods. In *Agile Data Warehousing Project Management Business Intelligence Systems Using Scrum*, Hughes provides detailed background information and guidance on iterative development, defining projects for iterative development, and adapting data warehouse projects for iterative development [7]. Specific to this paper, Hughes offers a six stage process to assist teams in adapting to an agile approach, and discusses best practices for scaling agile.

Table 1. Common Agile Themes

Approach	Iterative Requirements	Modularity/ Decomposition of Projects	Time Boxing of Capability Delivery	Team Structure	User Involvement
Golfarelli, Rizzi & Turrinchia	“Fact Prioritization”	“Incremental Process”	“Single Fact Iteration”	“Self Organizing”	“trained” and “shared language”
Dasgupta & Vankayala	“Market Driven” “Product User Survey” “Customer Product Analysis”	“Cohesive and Valuable Subsets” “Logical Components”	“Monthly Patch-Sets and Enhancement Releases”	“Schedules of onshore and Offshore Teams Overlapped”	“2 Product Experts... “Always Available”
Hughes	“80/20 Specifications”	“Homogenous work packages”	“Time-Boxed Iterations”	“Self Organizing Teams”	“Embedded Business

					Partner”
--	--	--	--	--	----------

CONCLUSIONS

As illustrated in the Table 1 – Common Themes, agile attributes are already common on DW/BI projects. This study limited the collection of attributes to the five themes; further investigation may reveal additional themes and attributes. As more DW/BI projects embrace and use agile methods, the data will continue to be compiled and highlight the inefficiencies associated with waterfall-like implementations. While there will always exist projects for which a traditional approach is appropriate, these will be far and few between. For some the ritual of large volumes of documentation, the illusion of every requirement defined upfront, and structured decision events will override the benefits of agile. For these people, this may signal the end of a career as they are simply not able to “tolerate crisis” and adapt to the new models that impact their professional lives [10, pp. 78-79]. However, those that embrace the revolution will have the opportunity to continue to shift the paradigm to greater innovation, collaboration, and efficiency.

REFERENCES

1. Agile Alliance. (2001). *Manifesto for agile software development*. Retrieved from <http://www.agilemanifesto.org/>
2. Dasgupta, S., & Vankayala, V. K. (2007). Developing Real Time Business Intelligence Systems the Agile Way. In *Systems Conference, 2007 1st Annual IEEE* (pp. 1-7). IEEE.
3. Ericson, J. (2006). A simple plan: interview with dm review editorial director jim ericson. *DM REVIEW*, 16, 18-21.
4. Futrell, R.T., Shafer, D.F., & Shafer, L.I. (2002). *Quality software project management*, Upper Saddle River, NJ: Prentice Hall PTR.
5. Golfarelli, M., Rizzi, S., & Turrinchia, E. (2011). Modern software engineering methodologies meet data warehouse design: 4WD. *Data Warehousing and Knowledge Discovery*, 66-79.
6. Hughes, R. (2012). *Agile data warehousing manifesto*. Retrieved from <http://agiledatawarehousingbook.com/agile-data-warehousing-manifesto/>
7. Hughes, R. (2013). *Agile Data Warehousing Project Management: Business Intelligence Systems Using Scrum*. Morgan Kaufmann.
8. Hunt, A. & Thomas, D. (1999). *The pragmatic programmer: From journeyman to master*. Boston: Addison-Wesley.
9. Kossiakoff, A. & Sweet, W.N. (2003). *Systems engineering: principles and practice*. Hoboken, NJ: John Wiley & Sons, Inc.
10. Kuhn, T.S. (1996). *The structure of scientific revolutions*, (3rd ed.). Chicago & London: The University of Chicago Press.
11. Schwaber, K. (October, 1995). Scrum development process. Workshop on Business Object Design and Implementation of the Object-Oriented Programming, Systems, Languages & Applications Conference, Austin, TX.
12. The Standish Group International. (1995). *The chaos report*. Retrieved from http://www.standishgroup.com/sample_research/chaos_1994_1.php
13. The Standish Group International. (2007). *The chaos report*. Retrieved from http://www.standishgroup.com/sample_research/chaos_2006_1.php
14. The Standish Group International. (2010). *The chaos report*. Retrieved from http://www.standishgroup.com/sample_research/chaos_2009_1.php
15. Turban, E., Sharda, R., Dursun, D., & King, D. (2010). *Business intelligence: A managerial approach* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.