Agile Analytics: How organizations can benefit from the agile methodology for smoother delivery of data-driven analytics projects

Masterarbeit

Zur Erlangung des Grades eines Masters of Science im Studiengang Webwissenschaft

vorgelegt von

Rafi Ullah

215202917

Erstgutachter: Prof. Dr. Harald von Kortzfleisch, Institut für Management

Zweitgutachterin: Denisse Mendoza Almanzar, M.Sc., Institut für Management

Betreuung: Denisse Mendoza Almanzar, M.Sc., Institut für Management

Koblenz, August 2019

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

	Ja	Nein
Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden.	☐	☐

ABSTRACT

Despite the inception of new technologies at a breakneck pace, many analytics projects fail mainly due to the use of incompatible development methodologies. As big data analytics projects are different from software development projects, the methodologies used in software development projects could not be applied in the same fashion to analytics projects. The traditional agile project management approaches to the projects do not consider the complexities involved in the analytics. In this thesis, the challenges involved in generalizing the application of agile methodologies will be evaluated, and some suitable agile frameworks which are more compatible with the analytics project will be explored and recommended. The standard practices and approaches which are currently applied in the industry for analytics projects will be discussed concerning enablers and success factors for agile adaption. In the end, after the comprehensive discussion and analysis of the problem and complexities, a framework will be recommended that copes best with the discussed challenges and complexities and is generally well suited for the most data-intensive analytics projects.

Keywords Agile Analytics, BI, Data Science, Data Analytics

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

- Business Analytics (BA)
- Data Science (DS)
- Systematic Literature Review (SLR)
- Extreme Programming (XP)
- Test-Driven Development (TDD)
- Product Owner (PO)
- Work In Progress (WIP)
- Feature Driven Development (FDD)
- Dynamic System Development Method (DSDM)
- Rapid Application Development (RAD)
- Data Warehouse (DW)
- Hadoop Distributed File System (HDFS)
- Value-Driven Development (VDD)
- Business Intelligence (BI)
- Big Data (BD)
- Big Data Management (BDM)
- CI/CD (Continuous Integration/Continuous Delivery)

1 Introduction

1.1 General

To be in competition, modern enterprises utilizing data for their competitive advantage. Organizations are collecting and curating data on a grand scale in support of innovative, data-intensive applications both to increase revenues and to improve operational efficiencies. It has become clear that the value of data lies in an organizations ability to operationalise it. To work on the ocean of data to create data pipeline and extract useful information for decision support Business Analytics (BA) is a great challenge (Valentine & Merchan, 2016). Moreover applying some data science models and get it to production as a working product is no less than a battle. Teams are not coding collaboratively; models have to wait for engineering and testing, which ultimately makes it difficult to manage the resources.

In data-driven digital transformation, the ability to derive and process data in a meaningful way offers a competitive advantage. It helps to create new questions over that analytics and identify new opportunities. Combining data with suitable agile methodologies makes the process of achieving business value very efficient and cost-effective.

Data Science, together with agile methodology, facilitates to observe more and beyond to solve the problems creatively and in a systematic way. Over time, agile principles and practices have been evolved, along with new challenges and future directions. New trends, including data science and fast analytics, have arisen as part of business intelligence (Dharmapal & Sikamani, 2016).

Analytics product delivery could not be accomplished through traditional waterfall software development model because analytics is directed more towards data discovery, getting the relevant information and then making sense out of it. The assumption is to how to utilize an agile methodology with a focus on knowledge discovery rather than software features development.

Agile Analytics is also based on a set of rules and guiding principles. Practically, It is a set of well-organized techniques, which can partially be customized for specific analytics project (Collier, 2011 p.7). Agile analytics comprises of all the practices ranging from project planning and management to monitoring. It also enables effective collaboration between all stakeholders, including customers, management and the teams responsible for project delivery.

1.2 Problem Statement

Ever since the introduction of agile methodologies in 2001. They had been playing a vital role in software development and has been applied excessively in many organizations ranging from start-ups to large enterprises. By applying agile methodologies, the organizations have delivered successful software development projects (Collier, 2011 p.9). Development of analytics projects including BI, big data and data science are not possible with a traditional agile approach. The focus of an analytics project is to create business value using data discovery. The agile methodologies will be applied to analytics software with a focus on information use rather than tool development. There is a growing requirement in the last five years to provide faster information and analytics for decision making (Collier, 2011 p.10).

In analytics, BI systems are provided to assist the decision-making for top management of an organization. For the success of any BI initiative, it is essential to deliver accurate information. To some extent, BI solutions still lag due to time constraints, which ultimately causes project failure. To tackle all these issues and to get the required information to management, BI system development should be agile.

With the emergence of data science, it became part of a much broader set of analytics applications across all industries. The common challenge of enterprise data science is that the specialized data science teams are working on the things which are not related to their expertise. With a traditional waterfall model of product development, it is challenging to create an environment of collaboration, data is mostly inaccessible, and data scientist are not aware of project ownership (Valentine & Merchan, 2016). With an agile approach to data science, it embraces collaboration and creativity. Changes can be made later as a project evolves, and stakeholder feedback is ongoing.

1.3 Research Questions and Objectives

In any organization development of analytics and BI system requires the people from much different expertise to collaborate to bring the requirements to the actual product.

RQ1: How have agile methodologies evolved with the emergence of data science?

RQ2: Which agile frameworks, methodologies and best practices are well known for data-intensive applications and how they could be applied?

RQ3: What are the enablers and success factors of Agile Analytics?

RQ4: What are the challenges organizations faced for adopting agile methodologies for analytics applications?

The objective of the research is to identify how organizations can employ the best agile practices to enhance productivity and successful completion of data analytics projects and create value for the organizations. The focus will be on how we can create value from data using Agile Analytics. It provides a way of exploring the data by focusing on finding the value (Tom, 2012). It will be done by understanding the factors that affect Agile Analytics. Any analytics application including BI, big data and data science systems comprises of different processes, and each process is composed of multiple steps. Then to be completely agile, all the steps and components should also be agile. At every step, all the components from BD, DS and BI and software development teams should work and collaborate smoothly (Collier, 2011 p.38). The bottleneck in any component will affect the project as a whole and will slow down the delivery. The outcome of this thesis would be the guidelines on how an organization can apply agile methodologies to data analytics projects.

1.4 Research Methods

Qualitative Research Methods

The essential purpose of the research is to deliver the factual knowledge by abolishing the myths and misconceptions which allow the people to act (Timulak, 2009). There are generally two methods for conducting research they are quantitative and qualitative methods (Yin, 2015). In this thesis, we will use qualitative methods. The systematic literature review (SLR) and document analysis are used as a research methodology. A systematic literature review (SLR) is an organized way to identify, evaluate and interpret research findings related to a specific research topic (Saltz & Shamshurin, 2016). It facilitates the identification of gaps and helps summarize existing research concerning process and methodologies used by agile teams to collaborate.

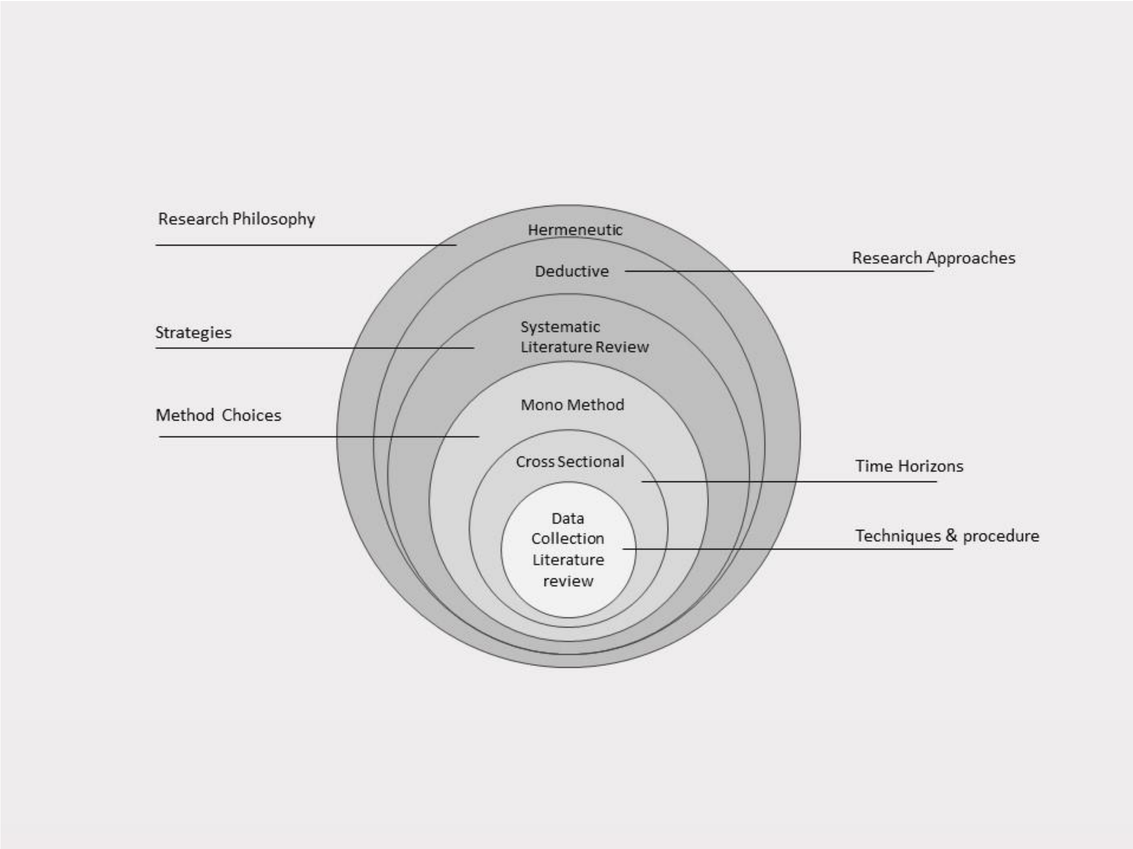

Figure 1 Research Onion (Own Image Following (Saunders & Schwaferts, n.d.)

2 Theoretical Background

2.1 Agile Development

The core idea and manifesto behind agile is "*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*" (Chang & Larson, 2016).

2.2 Agile Software Development

In recent years, agile methodologies have emerged as a whole new alternative to traditional software project management methodologies. There has been a consensus for a substitute for documentation driven software development processes (Ilieva, Ivanov, & Stefanova, 2004). The key challenges involved in software development are changes which occur and evolves at a much higher rate. In that case, if there occurs any change later in the project lifecycle, it will not only increase the cost but will also delay the project delivery. Agile methodologies have been introduced to counter these risks.

After the introduction of agile manifesto in 2001, many agile methods have been introduced, although these methods differ in particulars but serve the core idea of enabling the teams to accommodate the rapid change which ultimately reduces the project cost and make sure timely delivery (Paetsch et al., 2003).

Although every agile method is useful for a separate set of problems and contexts, the challenge remains how to determine which agile method is suitable for given project activities. All methodologies come with the risk and the main tasks for project managers is to identify and manage those risks by utilizing the most effective methodology for a specific situation.

2.3 Different Types of Agile Methods

The agile methodologies have emerged at a breakneck pace. There are multiple methodologies introduced for different problem contexts. We will discuss some of the agile methodologies and will compare how they are different in application and working.

2.3.1 Extreme Programming (XP)

Being the widely used agile methodology, XP supports the values shared by the agile manifesto for software development and also specify its own set of principles and guidelines (Paetsch et al., 2003). Extreme programming is based on a set of values. It is a way to work in coordination with corporate and personal values. It analyzes the progress and evaluates activities in the software development process. XP values include simplicity, communication, feedback, courage, and respect (Newkirk, 2002). Extreme programming is applied using a set of 12 practices. It works well mostly with a small development of up to 15 developers. Rather than focusing on a long process of requirements and design documentation XP focus on executable functionality and TDD. It emphasized robust regression testing, refactoring and code inspections through pair programming.

2.3.2 SCRUM

Scrum is an agile framework used for managing the entire lifecycle of software development projects. Sometimes discussed as an agile methodology, Scrum is a robust framework used for managing software development processes. It put extra responsibility to the teams to decide and come up with the working solution rather than providing detail for each step. The scrum model contains multiple components, but the roles, processes, and artefacts are considered as the main components. In Scrum planning meeting, teams commit to a specific set of the feature instead of task definitions. The scrum teams consist of five to ten people each, who work on the development of assigned features from the start to the end (Cervone, 2011). The teams are self-organizing and cross-functional. Overall there is no team leader to assign tasks, and all the issues and problems are solved as a whole team. As scrum is an iterative technique, it delivers the product features in an incremental way. Scrum could be applied to projects of either size from small, medium to large scale projects and even scalable to the whole organization (Livermore, 2007). Scrum contains two important roles Scrum Master and Product Owner (PO). Scrum Master’s responsibility is to make sure that the team is working according to scrum rules and guidelines. PO represents customers and conveys the business requirements and lead towards the required product (Angela Stringfellow, 2017). Scrum processes consist of five main activities, including sprint kick-off meeting, sprint planning, the sprint itself, daily scrum meeting, and at the end sprint review meeting (Cervone, 2011). Projects evolve with a series of sprints. In the planning meeting at start team members commit to certain features, and then sprint backlog is created to track the project progress.

2.3.3 Kanban

Organizations moving towards agility across the world are implementing Kanban to their existing software development processes to model their workflows in an agile way and faster delivery to market (Matharu, Mishra, Singh, & Upadhyay, 2015). Kanban is a lean method and popular framework used in agile software development. The robust feature provided by Kanban is to visualize the flow of work using charts with every column representing each step in the flow (Poppendieck & Cusumano, 2012). In software development, each work item is represented visually on a board called Kanban board, through which team members can see the actual progress and state of the work item. It is a method to create the product with continuous delivery without burdening the teams, which ultimately helps the teams to work together more efficiently. The Kanban offers visibility to each process in software development by showing assigned task to each developer and prioritize the critical tasks and spot the bottlenecks.

Moreover, its objective is to minimize WIP items and develop only those which are requested by the customer, which will produce the flow of continual delivery of items to the customer. According to team capacity, Kanban limits the work in progress items and balance the demand and output of the team’s delivered items. It helps to maintain a steady flow by visualizing the processes and minimizing the bugs (Ahmad, Markkula, & Oivo, 2013).

2.3.4 Crystal

Crystal is an agile development framework and a part of crystal methodologies family developed by Alistair Cockburn in early 1990s. Crystal, along with adaptive ideas, evolved together to provide guidelines for software development processes (Beck et al., 2001). Rather than tools and processes, it put more emphasis on individuals and collaboration between them, which will bring efficiency to the development project. Crystal is a very flexible model which could easily be integrated with other agile models and frameworks. The core idea is that there should be different treatment for every different project based on the situation. Crystal method is a methodology toolkit in which organizations combine elements from different methodologies to fit the individual project. Organizations develop different combinations of methodologies which fit their business needs (Livermore, 2007). It provides incremental development with not more than four months. To deal with time constraint an active communication between the teams and other stakeholders is recommended with support to lightweight documentation (Farhan, Tauseef, & Fahiem, 2009).

2.3.5 Feature Driven Development

Introduced in 1997, FDD is an agile development methodology which consists of two stages, first is to identify feature to develop and second is to develop that features iteratively. FDD can effectively control complex and incremental projects. It offers different capabilities to management including analysis and performs different tasks to add the value to the customer, checking the project progress with time and budget constraint, identify risks and ways to address and minimize them (Hunt, 2006). FDD consists of different phases which consist of

- **Development of the standard model**, the advanced and high-level model, and the scope of the system is defined
- **Build a feature list**, features are identified from information gathered after modelling
- **Plan by feature**, development plan for each identified feature is formed
- **Design by feature**, detailed design of feature is defined
- **Build by feature**, after detailed design teams start developing the feature

FDD can adapt to each project of a different nature (Firdaus, Ghani, & Yasin, 2013).

2.3.6 Dynamic Systems Development Method (DSDM)

DSDM is an end-to-end agile framework which not only addresses the whole project life cycle but also addresses its impact on business. Like other agile frameworks, DSDM put the focus on individuals rather than tools. It understands the needs of business and delivers the software as quickly as possible by providing best practices for rapid application development (Stapleton, 1999). Invented in 1994 while using Rapid Application Development (RAD) by project managers, DSDM represents and utilize much of existing knowledge about project management and became a famous framework among software development community for handling complex business projects (Voigt, 2004). It is known for its capability for rapidly delivering working software having budget and time constraints with changing requirements. It is adequate for a large scale and also on small solutions. The primary purpose of DSDM formation is to replace the unstructured software development using RAD. Being an iterative and incremental approach, it emphasizes continuous customer involvement. It has a total of four phases

Feasibility and Business Studies

In this phase, the technical feasibility of application is analyzed against the proposed problem. In addition applicability of DSDM is also verified for that specific problem. After carefully reviewing the business problem with system requirements, the underlying architecture of the system is defined

Functional Model Iteration

The prototype of the required system is developed iteratively for early user review and to get a clear understanding of the proposed system. The prototype is improved iteratively based on user feedback.

Design and Build Iteration

At this phase, the prototypes are tested to verify their usability in the operational environment. The components are refined further to achieve the required system.

Implementation

At this stage, the system is deployed, and the users are trained.

2.4 Business Intelligence (BI)

2.4.1 Introduction to BI

In the past two decades, BI has become increasingly crucial among business communities. BI is known as a set of standard processes, tools, and technologies needed to look insight into data for information and use that information to derive business plans and to make critical decisions. BI is an umbrella concept which includes big data and BA and knowledge management (H. Chen, Chiang, Storey, & Robinson, 2012). With the lowering cost to store and process a vast amount of data, there has been exceptional growth in the adoption of BI technologies from large enterprises in the last decade. Businesses are employing sophisticated data analysis techniques to make business decisions and to offer personalized services for customers (Chaudhuri, Dayal, & Narasayya, 2011).

With BI, companies can lower cost and improve business opportunities. It also helps to identify inefficient business process causing loss. Although BI comes up with huge promises, implementing a successful BI project is not less than a challenge. Data must be organized and processed before feeding the BI system to get better insights.

2.4.2 BI Architecture

The most essential in any BI architecture is the data warehouse and data marts on which any BI system can fully function. The classic BI architecture consists of multiple layers, including data sources, data warehouse, ETL layer, and end-user (Chaudhuri et al., 2011). DW is the central part at which data from multiple sources are gathered and stored after processing. The main purpose behind building a DW is that data must be stored centrally, which later could easily be accessible for analysis and decision making (Narasimhan, 2014). A data warehouse is considered as a critical factor for any successful BI implementation.

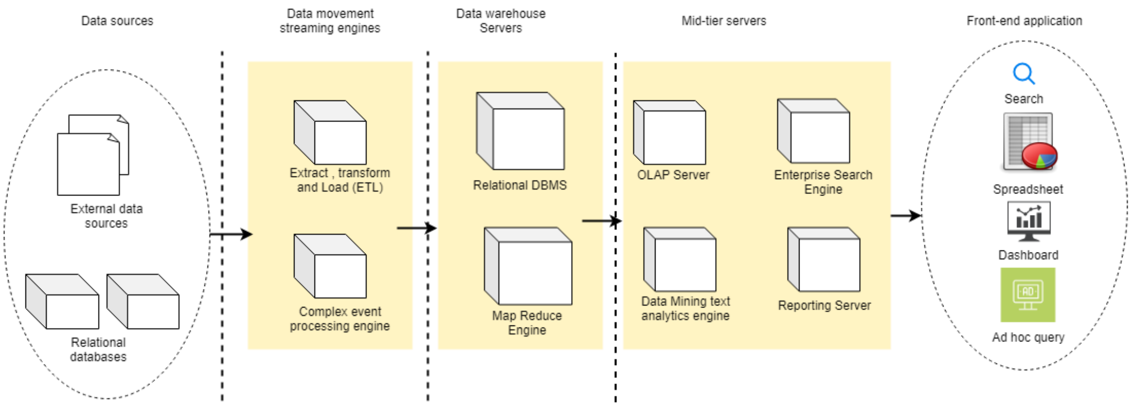

Figure 2 Business intelligence architecture (Own image following (Jasim Hadi, Hameed Shnain, Hadishaheed, & Haji Ahmad, 2015)

2.4.3 BI architecture Layers

BI architecture has the following layers

Data Source Layer

Data source layer comprises of several data sources which are operational database and external data sources. In a traditional setting, operational data refers to structured data stored in relational databases. The external data sources consist of social media and customer data (Jasim Hadi et al., 2015). The data from all these resources will be transferred to DW.

ETL Layer (Extraction, Transformation and Loading)

The ETL is the process to Extract, Transform and Load the data from multiple selected data sources to the data warehouse. In the extraction process data from external and internal sources are identified and extracted. The Extraction process is critical as it must be planned carefully as the later analytics process depends on this data. The organization must decide about the relevant data sources. Extracting all the data or irrelevant data will not be sufficient for decision making (Coelho Da Silva et al., 2018). The extracted data is moved to the staging area. The extracted data is different and is not useful for analysis. In the transformation process, data is cleaned, mapped and transformed by applying multiple operations before moving it to the data warehouse. The concluding step involves in the ETL process is loading. In this step, relatively massive datasets need to be loaded to the DW (Chaudhuri et al., 2011). Hence it needs optimization for better performance.

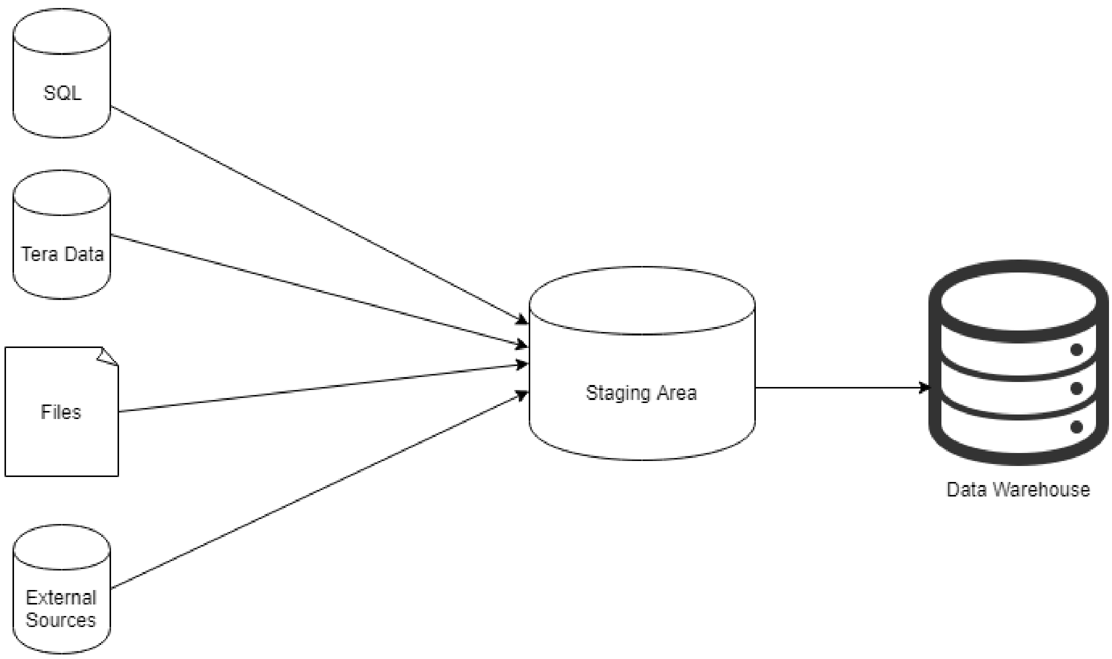

Figure 3 ETL process

Data Warehouse Layer

At this layer, all the data collected from multiple sources is loaded into DW. A DW is a storage location where data is stored to be directly consumed for reporting and analytics purposes. The data serves as a single point for serving different organizational analytics needs by preserving the data quality(Chaudhuri et al., 2011). In some cases, due to complexities and organizational requirements, data warehouses divided into subsets called data marts. Each data mart could serve the analytics requirements for different departments (Chaudhuri et al., 2011).

End-User layer

End-user consists of dashboards and analytics portals of an organization. The reporting tools display the data in a meaningful way. The end-user can manipulate and interact with the dashboard to perform analysis.

2.5 Big Data

The term Big Data is used for large scale datasets, both structured and unstructured, which are generated through day to day or other business transactions and activities. Diverse data is produced from several sources in different sizes. As the name implies, the term big data describes the data which cannot be captured and processed with traditional databases. Data from numerous sources, including the internet of things (IoT), social media and mobile devices data are the new sources of complex data. The Big Data sources include data from sensors, devices, logs, transactions, social, web and many other types of data from many different sources (Orton, 2015).

By analysing the big data, the researchers and analysts can find insights which could be used by business users to make decisions. Different analytics techniques such as machine learning, data science, natural language processing (NLP), and statistical analysis are applied on data to get in-depth insights.

Big data have many characteristics which differentiate it from transactional data that are Volume, Velocity, and Variety.

2.5.1 Three V's from Big Data

Volume

Volume denotes the amount of data. In addition to structured data, the big data systems have to process a large amount of unstructured data. Data could be a value from sensors, a social media feed, and log data, mobile, and customer touchpoint data or any unknown value (Swapnil & Anil, 2016). The size of the data can exceed to terabytes. Because of lowing the price or storage cost, it has become a standard to store the data for different purposes. Different technologies are available in the market to store huge amount of data, including Hadoop, which store the data in a distributed file system known as HDFS (Narasimhan, 2014).

Velocity

Velocity represents the speed at which data is created. As the flow of information is continuous, it created increasingly at a breakneck pace. To provide a quick response and to maximize efficiency, rapid processing is needed (Swapnil & Anil, 2016). In some processes, data flow must be analyzed as soon as it is created from the source. Ability to analyze real-time data provide actionable insights rapidly. The actual potential of data could only be unleashed based on how quickly the system can process the generated data. Velocity related challenges could be handled with new innovative data technologies like NoSQL database(Swapnil & Anil, 2016).

Variety

As discussed before, data is collected from several sources which could be structured, semi-structured and non-structured data sources. Structured data refers to relational databases and is stored in tables, semi-structured data refers to XML, JSON and the data from NoSQL, unstructured data consists of log files, text and multimedia content, sensors data and many different kinds of other business data (Narasimhan, 2014). Variety refers to the nature of data represented by various data sources with different data types. It is the biggest challenge to analyze a large volume of data of various types (Swapnil & Anil, 2016).

2.5.2 Big Data Tools and Technologies

Due to the size and types of data involves, big data requires the latest tools and technologies to process it. There are latest technologies available which not only provide the extension of traditional databases but also perform storage and processing of big data. Currently, Hadoop and Map-reduce are the leading tools to process big data. As a programming model map-reduce work on the Hadoop framework. It performs parallel processing by dividing the tasks into smaller sub-tasks (SINGH & SINGLA, 2015).

Hadoop

Hadoop is a distributed batch processing system used for parallel processing of big data. Hadoop has three main components, HDFS, YARN, and MapReduce. It is a scalable and distributed system which can work on minimum network bandwidth (Singh, Singh, Garg, & Mishra, 2015). Hadoop is very flexible and low cost to manage and process large data sets efficiently. It also does job scheduling for data and resources and provides load balancing (Coelho Da Silva et al., 2018). Hadoop Distributed File System (HDFS) is the most critical component of Hadoop to manage the data from different servers. It is file-based and does not require a data model to store data. It can handle all type of data from log files to XML and JSON files.

MapReduce

Developed by google MapReduce is the main technology which is implemented by google in an open-source environment. MapReduce permits to write optimized programs to process large chunks of data. The fundamental function of MapReduce is to divide a big task into several smaller chunks and treat them accordingly (Chaudhuri et al., 2011). MapReduce has a total of four components

Input The data collected to process is further divided into chunks which are allocated to mappers. These small chunks are called input.

Mapper Mappers are the smallest unit assigned for processing

Reducer Output from mappers serves as the input for the reducer which ultimately aggregates it for final output

Output All Jobs which are aggregated at reducer are collected here and represented as the combined output

Hadoop Distributed File System

HDFS is a data storage element of open source Apache Hadoop project. Able to store any form of data, HDFS runs on low-cost hardware and is highly scalable to thousands of systems. Based on master-slave architecture, HDFS is a file-based system. It converts large data files into standard 64 MB chunks and stores them in a massive cluster (Narasimhan, 2014). There are two nodes, data node and name node in each Hadoop cluster. The name node is considered as the master node. It controls all data nodes. Data nodes are regarded as slave nodes(Coelho Da Silva et al., 2018).

Hadoop Technologies	**Description**
Apache HBase	HBase is a NoSQL database which runs on the top of (Hadoop distributed file system) HDFS. Written in java, Hbase allows to read and write the data on the HDFS simultaneously.

Apache PIG	PIG is a scripting language. It is used to analyze large datasets. With PIG the developers can write the programs to process large dataset from Hadoop cluster. PIG convert the programs into map-reduce jobs and execute them. Initially, it was developed by yahoo for its data scientists.
Apache Sqoop	Sqoop was built to transfer bulk data between relational databases and Hadoop. It is used to import data between external data sources to HBASE, HDFS or HIVE, and vice versa. It allows parallel data transferring ability. It uses SQL queries and already saved data to perform import and export and to perform updates between Hadoop and relational database.
Apache Hive	Just like SQL, Hive is a query language called HiveQL. It was developed by Facebook, mainly for data analysis. It is adopted in many Big Data organizations. It also allows querying the data from HDFS, which later would be converted into map-reduce jobs.
Apache Flume	Flume assists in collecting and moving a large amount of data with the help of distributed services. With flume, the user can utilize log data and can perform parallel streaming to collect real-time log data from multiple sources.
Apache Zookeeper	Apache Zookeeper is an open-source project initiated by Apache. It provides centralized infrastructure to perform synchronization across the whole Hadoop

	cluster. It also provides naming services and configurations throughout the cluster.

Table 1 Apache technologies and description following (SINGH & SINGLA, 2015)

2.6 Agile Analytics

2.6.1 Introduction to Agile Analytics

The volume of data is increasing rapidly through both internal and external resources, and analytics tools are also getting more sophisticated to handle different domains of an enterprise. Organizations are trying to find value through analytics in the face of complex market competition. Decision-makers of any organization need rapid and precise analytics insights for making timely decisions to be ahead of competitors. Quick changes are required in the current system to cope with market threats, which ultimately end up in the organization with many parallel system developments as the companies are not prepared to handle the change at this pace. There is no unified and consistent approach to integrate new technologies with the existing systems (Earley, 2014).

Agile analytics provide flexible analysis and adjusting change to specific needs rather than following the rigid flow. Similar to agile methodology, Agile Analytics also comprises of a set of rules and guiding principles. By focusing on end goals, it provides data-driven predictions for decision making. It is a style for building DW, analytics, and BI application. It offers faster data discovery, which is well suited to changing business requirements. It focuses on iterative development lifecycle with an emphasis on finding value by analysing multiple data points. It reveals new interpretations and answers from result findings. It can analyze unorganized and raw data by prioritizing value over the set of processes. It will give the flexibility to examine the datasets with no real value and will also help to remove biases.

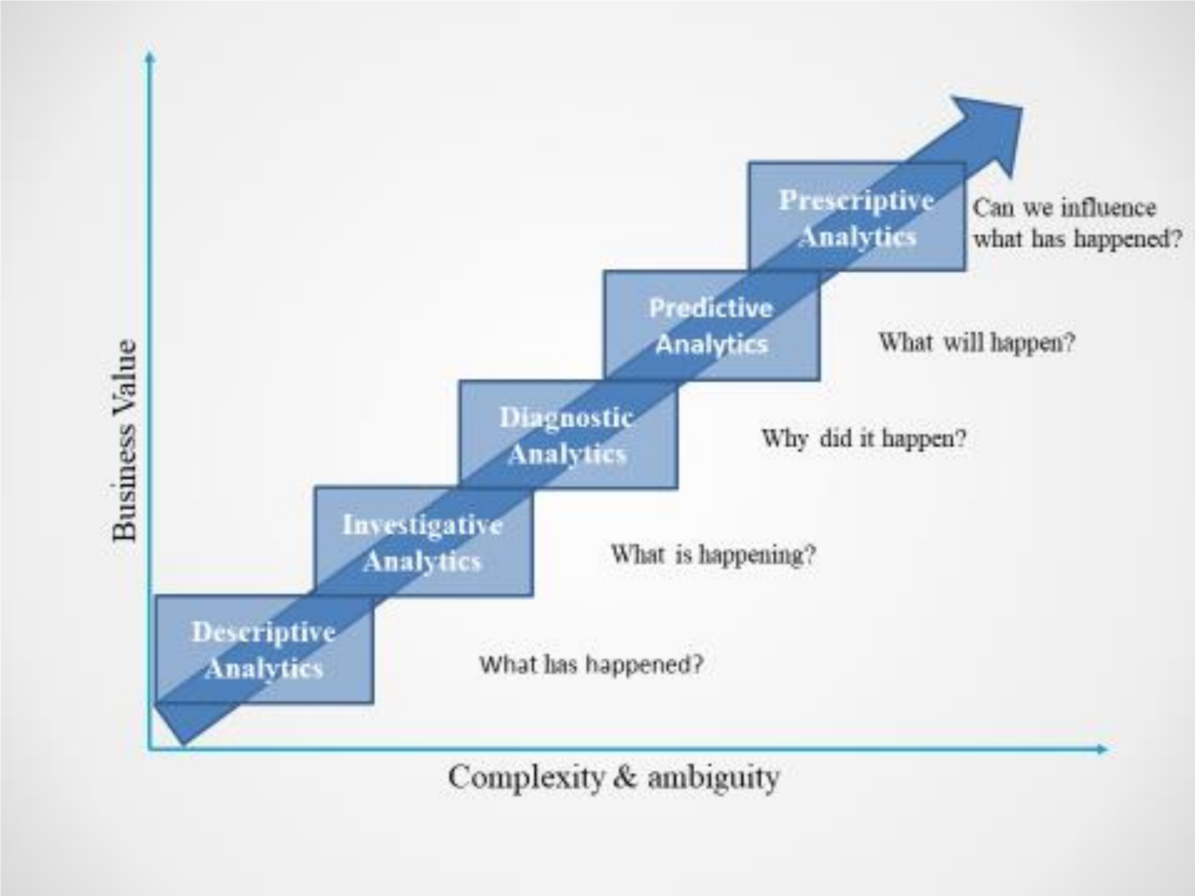

Figure 4.Analytics ladder (Own image following Tom, 2012)

Development of complex BI and analytics systems involves high risk, lengthy and complicated planning and massive uncertainty about project success. With analytics systems using the traditional system, a development method could result in a high-quality product. However, it requires much planning and is comparatively expensive to adopt, as it requires extensive design and documentation. Putting much effort before having even a small workable solution could be a massive setback for the company if the project fails. In that case, it is recommended to have only sufficient amount of planning, adoption to change, preparation for risks and involvement of team members to productively achieve the development targets. Often confused with methodology or framework, Agile Analytics is a development style. In every BI an analytics project, every team must support agile values and principles and adopt technical and team management and customer collaboration (Collier, 2011, p. 3-6).

Although Agile Analytics delivers meaningful information at a swift pace, it requires some changes in the organizational approach to handling the data. With technology advancement, it has become easy for businesses to identify useful information from data.

Agile analytics can accommodate changes and can scales very fast upon new requirements and impacts. It does not require long project timeline and huge investments to start (Tom, 2012). Tasks from business use cases related explicitly to analytics are assigned to cross-functional teams. For specific tasks related to another specialized area, a member from other teams could also be injected in the analytics team temporarily.

With the main focus on high-quality BI and analytics solution, Agile Analytics have multiple goals as explained below

Iterative and Incremental Development

Although it is considered as a modern approach, iterative and incremental development dates back to mid-1950s. As the development world embraced agile methodologies, incremental, iterative and evolutionary style of development widely applied, replacing the traditional waterfall model. It offers incremental development of a system by prioritizing user-valued features and evolve the system by continuous adaption of customer feedback (Larman & Basili, 2003). Work is performed in small iterations from one to three week and not more than four weeks. These small iterations with continuous user feedback will keep the project in the right way.

Value-Driven Development

It is generally believed that goal-oriented development to create business value could more effectively be achieved using value-driven development. The aim is to get a value feature at the end of every iteration (Ktata & Lévesque, n.d.). In VDD, application development is directly associated with business economics values. Irrespective of complexities and steps involved in the development of data infrastructure, users are only concerned about the outcome which ultimately would solve the business problem (Souza, Moreira, Abrahão, Araújo, & Insfran, 2016).

Production Quality

After each iteration, the provided feature must perform as per business requirement. It must be tested and debugged accurately at every stage of development. As the product evolves with every iteration, the outcome should be a working functionality. Continuous testing should be planned and performed rigorously to get a quality feature. The feature should be accepted finally after getting approval from the end-user.

Automate Routines

Agile development discourages to perform routine processes manually. To be fully agile, routine processes must be automated. Manual testing should be replaced by automated testing. Manually testing every test case causes valuable time loss. Automated testing will support to validate the system repeatedly to ensure product quality. With the emergence of DevOps, Continues Integration and Continuous Delivery could be performed, which will reduce considerable time to production. Analytics teams automate any repeated process and can put more focus on developing quality features.

Team Collaboration

The analytics project requires expertise from different areas comprising software development to data engineering and data science. Collaborations among all stakeholders are crucial to project success. Establishing a workspace which provides a collaborative environment and promotes positive collaboration among team members (Hossain, Ali Babar, & Paik, 2009). Continuous interaction and collaboration between the user and the development team with a daily meeting among technical team members ensure a successful project.

Self-Organization and Self-Managing Team

Create self-organizing independent teams which can manage their tasks autonomously. Hire the experts and provide them with the required tools, then allow them to be successful on their own. Agile project management enables team members to work independently and facilitate collaboration between end-user and other stakeholders in the project. The teams will decide the features to deliver in every iteration and will be fully responsible for committed modules.

2.6.2 Phases of Agile Analytics

Feasibility and Planning Phase

Being the first phase, it includes user requirements collection, identifying stakeholders and develop product backlog. Customer expectations are being set after a deep understanding of business and data. User stories which also serve as product backlog are written and explained by the product owner. Features are planned to develop and deliver independent from each other. Different tools and technologies are assessed and identified to be used in the analytics phase. Different data sources are also identified to collect data and infrastructure is also planned. Resource identification and allocation, along with the release iteration plan, are prepared (Dharmapal & Sikamani, 2016). Cross-functional self-dependent teams are made and assigned tasks for iterative development of identified features. Despite core team members, other members could be added later as the product evolves.

Implementation Phase

Features are prioritized and developed iteratively from product backlog prepared in the planning phase. Also, data is collected from several data sources that will be transformed into the standard scope and combined to central storage. Data cleaning is performed by removing isolated and unrelated data. At this phase daily stand up meeting is performed to increase collaboration between all stakeholders including data analyst, business owners, development team and product. Collaboration among teams is essential for successful and timely product development (Dharmapal & Sikamani, 2016). This phase also allows parallel development of user interfaces (UI) for data visualization. Pre-deployment model testing is performed to ensure that the newly developed model is compatible with the existing system and interfaces. One of the critical features for Agile Analytics framework is that it facilitates the parallel development of application components including UI, Model development and visualization. After the pre-deployment pilot phase, the results are evaluated from the management to assess the business impact. If the outcomes fulfil expectations, the project will move to the next stage, and if it deviates, the modifications will be performed. The project could also be abandoned if the required output is not possible (Tom, 2012). Once the project passes the inspection, it will be moved to the production environment. At this stage, the project has deliverable consisting of new dashboard, reports and analytics solution (MUNTEAN & SURCEL, 2013).

One of the principal value of agile development is that it can accommodate changing business requirements, and the final deliverable would be a required product. Predictive analytics could also be performed on data using analytics tools. Continuous integration and testing for different environments in every iteration will ensure the required project without any deviation (Dharmapal & Sikamani, 2016).

Final Phase

The final phase makes sure that all the requirements have been successfully implemented. It is the final stage before the product is released. Final testing is performed, and models are evaluated. The output product from this phase will benefit to end-user to make business decisions.

2.7 Data Value Chain for Analytics

Data value supply chain describes the data flow process in an organization and explains the essential tasks require to perform in each task to acquire valuable business value. It has been used as a process model for the development of a decision support system. It consists of multiple activities that need to be performed in a sequence to get valuable data insights. In the information value chain, organizations employ multiple different activities (Curry, 2016). A systematic review of each activity would be helpful to provide insight into value generation and hurdles involved in that process.

The data value chain explains the insight into the data evolution, starting from collecting raw data to perform analysis until the final predictions for decision making. The framework to generate the value consists of four main activities by which organizations create the commercial value of their data.

1. Data Discovery
2. Integration
3. Data Analysis & Delivery
4. Data Governance

2.7.1 Data Discovery & Acquisition

Every organization needs data sources to perform analyzes and make predictions; Data discovery involves the collection, preparing and organizing the data sets (Miller & Mork, 2013). With the emergence of the latest data technologies, it has become possible to collect, store and process a massive amount of data without having full information about its structure and meaning. Hadoop is a commonly used software invented by google to land a massive amount of data. It manages data storage and processing in a distributed environment without any need to do data modelling (Chang & Larson, 2016). Considered as a hub of data initiatives, including predictive analytics, machine learning, and data mining application, Hadoop supports structured and unstructured data and gives the user more flexibility.

The first step of the Data Value Chain is to collect the data. It could be from multiple sources, consisting of both internal and external. Data could either be human or sensors generated. In the case of social media when a user posts something, a range of data is generated including user location, device information, users like and dislikes and content of data. Considering the total number of that specific social media platform, the total amount of data collected is beyond imagination. In other forms, all B2b transactions, financial transaction, flight, and aviation systems are some of the other few examples to understand the amount and significance of data generated through customer touchpoints daily (*The Data Value Chain*, 2018). With the rise in the popularity and adoption of the Internet of Things (IoT) and industry 4.0 digitalization, sensors and recording devices also create a massive sum of data.

The first step is to create the inventory of available data sources and to evaluate the quality against different parameters. The crucial step is to convert unstructured data into structured data (Miller & Mork, 2013). After identifying and collecting the data, in the next step, it is filtered and cleaned before putting it into the DW. Analytics is performed over the data from DW. Data discovery is one of the biggest challenges in the whole analytics process as the later steps fully dependent on this (Curry, 2016).

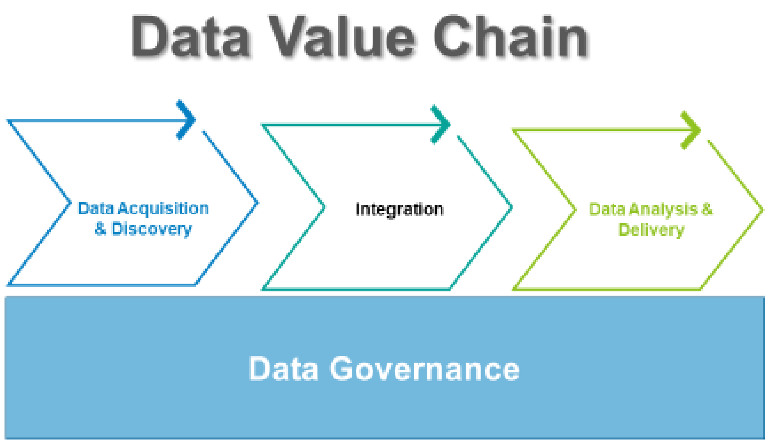

Figure 5 (Own Image following Nagle & Sammon, 2013)

2.7.2 Integration

After deciding the data sources and performing data acquisition, the data which is collected from multiple sources would be transformed into the common representation which is appropriate for performing analysis. Mapping is performed to relate the data from multiple sources. Physical integration is performed in the form of DW. It is also possible to integrate virtually through federated models. With the latest semantic web technologies, it has become possible to query on combined data in a non-tabular form. Relational databases are more suitable for tabular data (Miller & Mork, 2013). As the data is integrated in a structured way, its value increased. Data is stored in a convenient way for fast access to quickly perform analysis to get the strategically useful information (*The Data Value Chain*, 2018).

2.7.3 Data Analysis & Delivery

At this stage, data is analyzed to get data insights for decision making and delivered to the customer. It involves exploring the data to extract hidden information which could make an impact according to the business needs. There are many different methods of analysis, the simple and basic approach to define mathematical and statistical models (Curry, 2016). There are many sophisticated methods in practice which are currently being used to get insight into data. The results from this stage would be used to make decisions. To better understand the problem, visualization is performed in the form of static result report or application with interactive charts and graphs. By visualization, end-users can easily understand the meaningful information to quickly grasp the problem and make decisions (Miller & Mork, 2013). The final aim of the whole process is what actions should be taken after visualizing the problem. If the visualized results are showing the trends in loss, the stakeholders can take necessary steps and make a decision which could positively affect the situation to convert the loss into reward. The end-user should be able to grasp the information and interact with the application without any difficulty(Curry, 2016).

2.7.4 Data Governance

Data Governance is a collection of rules, policies, practices, and processes which ensure the data management during the whole analytics process. It deals with the availability, integrity, and security of data to make sure quality data is available for analysis at every instance of analytics implementation phases. A Data Governance program consists of a set of rules and procedures and a governing body to plan and execute those procedures (Tallon, 2013). Considered as one of the main force for driving business value from data, Data Governance is a success factor for any analytics initiative. It has a direct link to data quality, compliance, and security of data.

2.8 Business Intelligence and Analytics

Increasingly available robust technologies to process BD, BI, and advanced analytics have become very significant. The prospect linked with data analytics for different organizations has helped to evolve techniques, technologies, practices, and methodologies to analyze the business data to assist the business to understand market opportunities and threats and to make the timely decisions. In addition to utilizing tools and technologies, business analytics provides business-centric approach and methodologies which have a productive impact on applications from various industries including e-commerce, healthcare, security, financial institutions and more (Lim & Chen, 2013).

The terms Analytics, Business Analytics, Advanced Analytics, and Business Intelligence are related to one another. Analytics can be well-defined as the process or practice which analyzes, explore, visualize and find trends in data by utilizing information system, statistical and operations research methodologies. Analytics is a common terminology equally applied to all disciplines (Schniederjans, Schniederjans, & Starkey, 2014). It is the ability to collect, manage and analyze massive datasets to get rightful information at the fast pace and to present it to the management. It consists of not only tools and technologies but the methodologies specific to this process. Some organizations have adopted the latest technique and technologies in BI systems. Applying new techniques like predictive modelling and interactive visualization in analytics is called advanced analytics (Halper, 2017). Traditional BI systems provide transaction and cost management activities with visual representations in the form of dashboards. It includes ETL, DW and OLAP functionalities. With self-service and advanced data analytics techniques, users can interact with user interfaces, perform predictive analytics, visual data discovery, reporting, and data analysis in a more advanced and sophisticated way (Halper, 2017).

Sometimes referred to as Business Analytics, Advanced analytics offers a range of algorithms to perform complex analysis on structured and unstructured bulk of data. It uses sophisticated machine learning models, statistical formulas and other cutting-edge tools and technologies to find patterns and behaviour of data to make predictions and to optimize decisions. It is not only an algorithmic process but a complete set of activities consisting of the infrastructure of data management and processing, organizational processes to adopt data-driven environment, and development techniques (Halper, 2017).

2.8.1 Levels of Business Analytics

There are many levels of analytics, but the three primary levels are discussed below

1. Descriptive Analytics
2. Predictive Analytics
3. Prescriptive Analytics

Descriptive Analytics

In the analytics process, the first step is to identify and reason about past events about what has happened and why? Descriptive Analytics answer these questions (Mohr & Hürtgen, 2018). Performed on big data, descriptive analytics help to discover hidden information, and explain the characteristics and relationship among the characteristics. It provides brief information about what has happened and why and what is happening right now, e.g. pay-per-click and email marketing (Sun, Sun, & Strang, 2018).

Predictive Analytics

Used as a recommendation, Predictive Analytics describes the events which could happen in future. Predictive analytics algorithms such as machine learning, regression analysis, neural networks exist since long. Though, the infrastructure and technology advancements in recent years have made it very convenient to use these algorithms in analytics application. Predictive analytics have been widely used in the marketing industry to better understand customer needs and preferences (H. Chen et al., 2012).

Prescriptive Analytics

In prescriptive analytics, analysis is performed on raw data, and the best solution is recommended. It reasons the information about expected possible scenarios, past performance, available resources, and suggests the strategy for future actions. It combines and utilizes both descriptive and predictive analytics (Sun et al., 2018). Based on the results from predictive analytics, it suggests all the favourable outcomes and proposes a future course of action to bring in a particular outcome. It uses feedback and learning system to improve the relations between suggested action and required outcome. Recommendation systems used by Spotify and Netflix are well-known examples of prescriptive analytics.

2.9 Past and Existing Agile Analytics Development

Ever since the introduction of agile development strategies in software development. They have been proved very significant. However, there has not been satisfactory research done on how to use agile methodologies for the development of an analytics application. Development of analytics systems such as BI system is incremental and always needs some revisions to implement changing business requirements (Ko & Abdullaev, 2007). Agile methodologies can manage this environment. As being specific to the organizational needs, there is no standardized approach to apply agile methodologies.

There are some proposed agile methodologies for the development of individual stages of the analytics process, but no particular methods for the development of complete analytics solutions or incorporating change in the existing one (Knabke & Olbrich, n.d.).In the book "Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing", ken collier in 2011 introduced and implemented agile methodologies that specifically deals with the development of Data-Intensive solutions(Collier, 2011).

2.10 Agile BI

Agile meaning is the ability to be flexible to adopt change. Defined by Forrester Research Agile BI is a methodology which combines all stakeholders including business processes, organizational structure, tools and technologies to assist the decision-makers (Evelson, 2014). There is an abrupt shift towards agile methodologies from the traditional waterfall approach. The popularity of agile methodologies gathered support and attention of people of the data community (Kannan & Services, 2011). Adoption of fast analytics, particularly machine learning and data science, natural language processing, predictive analytics, and artificial intelligence, are not supported in traditional BI systems. Traditional BI system mostly provides static views on structured data, which contradicts the agile manifesto which allows to adopt changing business requirements. There are some disadvantages of traditional BI systems mentioned below in the table

Disadvantage	**Problem**
Massive duplication of data	With every change, there will be a new requirement to change duplicate data,

	which causes data inconsistencies and will reduce data quality.
Different tools required to perform different tasks	As the meta specifications are not shared across tools, different tools are used to perform a different task which in turn create data inconsistencies
Strong relational data models	Due to structured relational data models, it is very inflexible to make any change. Analysis of external and unstructured data is also minimal
Waterfall model approach	With the waterfall approach, there are very long development cycles with less user involvement. Very inflexible to adopt changing analytics requirements. Testing is performed at the end.

Table 2 Disadvantages of traditional Business Intelligence Approach following (MUNTEAN & SURCEL, 2013)

For BI projects, teams must be collaborative and well organized. The BI systems should be developed with a focus on creating business value rather than getting more into technological details. Typical BI systems consist of three layers consisting of ETL, DW and Front End (Ko & Abdullaev, 2007). Organizations have multiple data sources from internal to external data sources with many different types of data, including databases, files, and emails, excel sheets, audio, and video data and more. Traditional BI system does not utilize all the data but instead, rely on small fractions for the required result. Some BI systems only depend on structured data (MUNTEAN & SURCEL, 2013).

The BI projects should be planned iteratively rather than provide all in one package. Initially, small but working functionality would lay the foundation of later effective delivery until the maturity and successful product delivery (MUNTEAN & SURCEL, 2013).

Most of agile BI project teams use Just-In-Time planning with iterative and incremental planning to accommodate rapid change in the requirements. Customers provide continuous feedback and work alongside the development teams to keep development in line with business needs. Teams deliver partial but fully functional features after each iteration. The product evolves with every iteration to accommodate all changes. Teams and customers collaborate to accommodate changes and to work together to achieve the desired product (Leybourn, 2013). Customers expect fast delivery without waiting for months. Agile methodologies can make a positive impact on users by delivering value information quickly rather than waiting for months. The approach of applying agile methodologies on BI systems would be how to get the required information rather than building a software application (Chang & Larson, 2016). Agile BI provides timely information in a proper format to the right person.

Application of agile in BI is still not much researched upon in academia, but it is a buzz word and highly discussed in the business world. The most essential and highly applied aspect of agile is agile software development. Although software development using agile methodologies is widely known, agile business intelligence is a less known but a vital aspect. There are some success factor for agile BI discussed below

2.10.1 Factor Effecting Agile BI

Iterative Development Life Cycle

In agile software development, the whole process is iterative and incremental. Rather than spending months and years to come up with finished product consisting of all required functionalities, the agile development put more focus on rapid delivery of partial functionalities get the customer feedback to refine the features for the next iteration. This iterative and incremental approach also well suited for large scale analytics projects. In BI implementation software is the constant, while the other things include creating the data connections between software and data sources and creating the user experience are variable. As the user experience is consists of a large part of BI systems, profound user involvement is critical in successful implementation (Intelligence, n.d.). Teams are working together in an iterative process to extract and derive useful knowledge in each iteration. In BI, each iteration is based on user stories, and the end of each iteration, the release represents each story (Intelligence, n.d.). The whole systems are divided and developed in small chunks. The primary purpose of iterative development is to get customer feedback and to validate the project progress, which means is that every iteration should provide working functionality. One of the critical task to apply agile methodologies to analytics projects is to adopt a personalised project management process which supports iterative and incremental approach (Collier, 2011).

Flexible to Adopt Changes

The development process should flexible to adopt changes as the product evolves. The organizations need to define the strategy to predict and cope with the sudden changes. The changes could be different types depending on nature. Internal business changes could be a change in strategy or reorganization (Strohmaier & Lindstaedt, 2005). External changes could be the emergence of latest technologies and change in demand dimension. Agile analytics is to identify and respond to changes quickly. The system must be able to predict future changes by using collected data and should respond rapidly (Woolley, Melbourne, & Hobbs, 2008). BI applications are developed to answer the strategic question according to changing and new business needs. Agile BI could be viewed as the ability of a business to provide valuable information in response to rapid changes. In advanced business analytics solutions, the ability to realize the upcoming changes by performing data comparison will provide the ability for early adoption (Sanaa, Afifi, & Darwish, 2016).

Automate ongoing BI Process

Foundation of creating the analytics application utilizing agile methodology depends on its ability to automate the repetitive task and to focus on quick insights into data to derive fast decisions (Valentine & Merchan, 2016). The routine manual process should be automated, and teams should implement automated continuous implementation and deployment. With cutting edge latest technologies like machine learning and artificial intelligence, and data science systems are capable of automating as much workflow as possible. Increased automatization and intelligence of BI systems will increase the efficiency to analyze in a short time by reducing manual inputs (Sanaa et al., 2016). Time for each iteration could be minimized by automating testing and deployment (H. M. Chen, Kazman, & Haziyev, 2016).

Team and User Collaboration

Collaboration among teams, team members and customers are pivotal to successful implementation. As compared to traditional approaches, agile put more focus on collaboration, which in turn enhance productivity. In an agile setting, customers work alongside development teams until project maturity (Kruchten & Gorans, 2014). Collaboration with the customer is valued over contract negotiation. Changes made alongside development iteration could be reviewed quickly to validate customer expectations. Without customer feedback, functional documents will define how the product should look like; these documents sometimes deviate from customer expectations. With the agile approach, customers can see the working solutions throughout the whole project (H. M. Chen et al., 2016).

Technology/ Technical Factors

Applying agile methodologies to the BI system will change the way the infrastructure was being used for testing, build, and deployment from many different teams. The induction of new Agile Development and Operations (DevOps) tools that optimize and reduce the time to delivery and support frequent releases to end-users. Tools and technology support is essential for any agile team to be critical (Kruchten & Gorans, 2014). Although considered as the most crucial factor in BI development, all other things should be in place before the technology phase. Data for business intelligence is collected from multiple sources; the sources could be structured, semi-structured or unstructured. Its ranges from spreadsheets, logs and sensors data, JSON files and databases. Data representation is different for each source, and some sources will have data quality issues, which later could affect the development process and lead to an unsuccessful implementation (H. Chen et al., 2012). To tackle this problem, organizations create an integrated data source to make it consistent with using with BI and analytics system. Every organization has a different approach for data integration, which ultimately decides project success (Williams & Williams, 2007). Considered as the most critical and integral to BI, the DW integrate and serve the data to BI systems, and the way teams handle it affect the agility in BI systems (Baars & Zimmer, 2013). One of the significant components of the analytics system is frond-end application. The business value created through processing on extracted data could only be understandable by the user if displayed correctly at front-end (Nagle & Sammon, 2013).

Front-end technologies are directly accessed by users, hence critical to Agile BI implementation. To be agile in BI implementation, the whole components of this process should be agile (Krawatzeck, Dinter, & Thi, 2015). Agility in front-end application could be achieved by introducing new tools or by modifying the existing ones as needed. There could be multiple different tools involved in this process (Baars & Zimmer, 2013).

Alignment of Goals and Business Value

Goals and objective for implementing Agile Analytics must be cleared. Every business has different reasons and objectives; some adopt it to increase productivity, to improve customer involvement, manage project risk and to address the quality problem and more.

There could be plenty of reasons to shift to an agile way. The goals must be cleared and evaluated against the success factor. The organization should also analyze how the objective supports business goals. It will ensure achieving a strategic objective by enabling the improvement in analytics processes (Collier, 2011 p.302).

Training and Coaching

Formalized way of transferring knowledge is represented as training. Agile training help to understand agile practices, principles, and values to newly formed Agile groups. Sometimes referred to as a scrum master, the agile coach is a vital member of an agile community who has deep understandings of agile techniques. The coach could be a mentor, collaboration conductor, and problem-solver. Successfully adopting agile methodologies require a focus on coaching and training. All team members, including managers, need agile training (Collier, 2011).

3 Literature Review

3.1 Introduction to Literature Review

The literature review is the way of exploring by targeting a specific area of research to define the approaches to answer the questions. Focusing and elaborating the work which has been done on a specific topic, it may give the contextual knowledge to perform higher work or to answer at its own(Levy & Ellis, 2006). It analyzes and combines information from scholarly literature about significant issues by comparing and evaluating main arguments, methodologies and approaches.

According to (Webster & Watson, 2002), the literature review is the process of reviewing the existing literature to create a foundation for proceeding knowledge. It assists in the development of theory and to uncover areas where more research efforts are needed. It could be a necessary part of a research report, article or dissertation and could also be used as a standalone research paper.

The literature review must be structured to facilitate the reader to grasp the idea entirely. Following are the “five C’s” of writing literature review discussed in the table below

Cite	The researcher should always keep the focus specific to the research.
Compare	Compare the arguments and theories presented in the research paper or literature to get the authors finding and agreements. Find out who has employed the same approach?
Contrast	Contrast several arguments, methodologies, and approaches explained in the literature to find the point of disagreement and debate.
Critique	Find out the persuasive arguments and the fact behind its persuasiveness. Which findings are the most reliable ones and why? Put more attention to the assertive verbs used by the author.
Connect	Connect the findings with the area of research and draw perspective based on what has been said in the literature.

Table 3 "five C's" for literature review following (Labaree, n.d.)

The difference of literature review to the book review and annotated bibliography

Annotated Bibliography	Book Review	Literature Review
Annotated Bibliography reviews and aggregate the relevant sources and explain its significance with relevance to the research question.	It involves analysing and evaluating the book.	A literature review involves surveying all relevant literature to identify already revealed and hidden facts about a particular topic.

Table 4 Difference to Literature Review

3.1.1 The need for conducting a Literature Review

Form a concrete theoretical foundation for research

One reason behind conducting the literature review is to analyze and explore the research and outcomes which has already been done regarding the specific area of interest. It lay the solid foundation to perform quality research based on analysis and synthesis of available literature. Although all information from literature is not equally accurate, the researchers need to use quality literature as a basis for further studies. It ensures the legitimacy of research performed and brings out reliable findings (Levy & Ellis, 2006). By providing the foundation for a proposed area of research, it also authenticates the existence of the research problem (Abbott & Grady, 2011).

Conducting a useful literature review not only provides a theoretical foundation for further studies but also help to choose the precise research methodology for further studies. It should also provide the researcher to justify the selection of specific methodology as to why applying that approach is more suitable in that specific area of research (Webster & Watson, 2002).

How and which literature fits with the research topic and vice versa

Based on the concept-centric approach, to conduct a useful literature review, the researcher must continuously validate how the following literature is related to his area of research. The answer to this question will keep the researcher on track to related literature. Furthermore, the researcher should focus on the literature, which corroborates with the problem related to the research area on which the researcher is working. It will give the solid ground to the researcher to perform research on that problem (Levy & Ellis, 2006). Not only should the researcher try to make the literature fit into his study but also their study should fit into the literature body of knowledge.

3.1.2 Stages of Literature Review

The simplest version of conducting a literature review can be explained in three steps, 1) Input, 2) Processing, and 3) Output. In this, the main and the most critical is the processing. It involves performing multiple activities sequentially, including collect, comprehend, analyze and synthesize. The output of the literature review should contribute something new to the body of knowledge. Strictly following the three steps guarantees the effective literature review for a beginner researcher (Levy & Ellis, 2006).

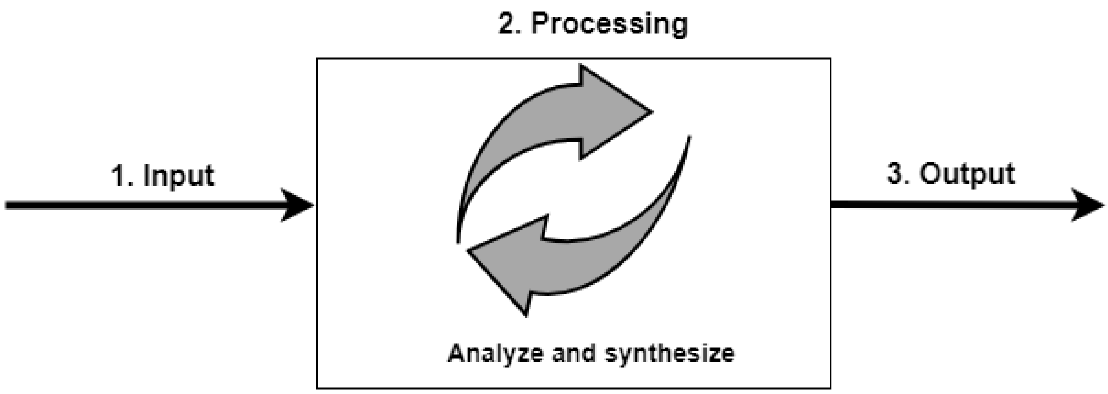

Figure 6 Stages of Literature (Own image following (Levy & Ellis, 2006)*)*

3.1.3 Structure of the Literature Review

The structure of the literature review consists of three main parts

1. Introduction
2. Body
3. Conclusion

Introduction

The introduction describes the topic with some elaboration about its significance and a statement about the conclusion that the researcher will extract after analysis and synthesizing the review. Moreover, in this, the researcher also discusses the importance of research related to the research question (Skene, n.d.).

Body

In this section, the researcher examines the past research with a focus on agreements, arguments, and methodologies. With the focus on the area of interest, the researcher will find the gaps and will state how his work will respond to it (Andersson, Beveridge, & Singh, 2015). Each paragraph in the body should deal with the separate theme of the research topic. Several reviews are synthesized in each paragraph to make a clear connection between different sources. Each source must be critically analyzed to contribute to filling the research gap (Willans, n.d.).

Conclusion

A conclusion provides the overall summary and findings of the literature review (Andersson et al., 2015). It explains in detail the overall literature and which literature leads to conclusive findings. It provides the answer to the research questions and gives suggestions about future work in this direction (Willans, n.d.).

3.2 Selected Research and Literature Review Methodologies

3.2.1 A Short Guide to Action Research (Johnson, 2012)

Action research is proposed by Kurt Lewin in 1940 in the United States. Using it changed the way the researcher interacts with the research setting. The name "action research" represents its working, where the work is not separated from investigation to conclude and propose the actions to solve the problem. According to (Johnson, 2012), action research deals with a variety of investigative and analytic research to identify problems and weakness. It involves studying a real school situation to improve the quality of direct practices. It is a systematic and sequential way to observe the practices applied, their problems and recommend possible treatment to solve that particular problem.

The action research consists of five essential steps

1) **Identify a problem** or topic area to perform research to determine the course of studies.
2) **Decide which data should be collected**, identify data sources and decide the frequency of data collection.
3) **Collect and analyze data** from sources identified in the previous step.
4) **Describe findings and how they could be applied,** create a plan to apply a course of action based on the finding from the analysis step.

5) **Share findings and plan of action** with others

Action research does not progress in a linear sequential way always. Based on the situation, some or several processes need to repeat to get to the required results.

Although there is much freedom to collect, analyze and present the findings, action research is still a very systematic approach. It contributes to finding the answer to questions which are unknown. It is not a very complicated method and is considered as very problem-focused, precise and accurate methodology. To adequately apply this methodology, an in-detail study must be planned before starting to collect data.

The length and time duration for every action research varies. It depends on how big the problem is and the time needed for data collection and analysis. Observations should be at regular intervals independent of its duration. By performing a detailed literature review and relating the conclusions to it delivers a contextual understanding of research.

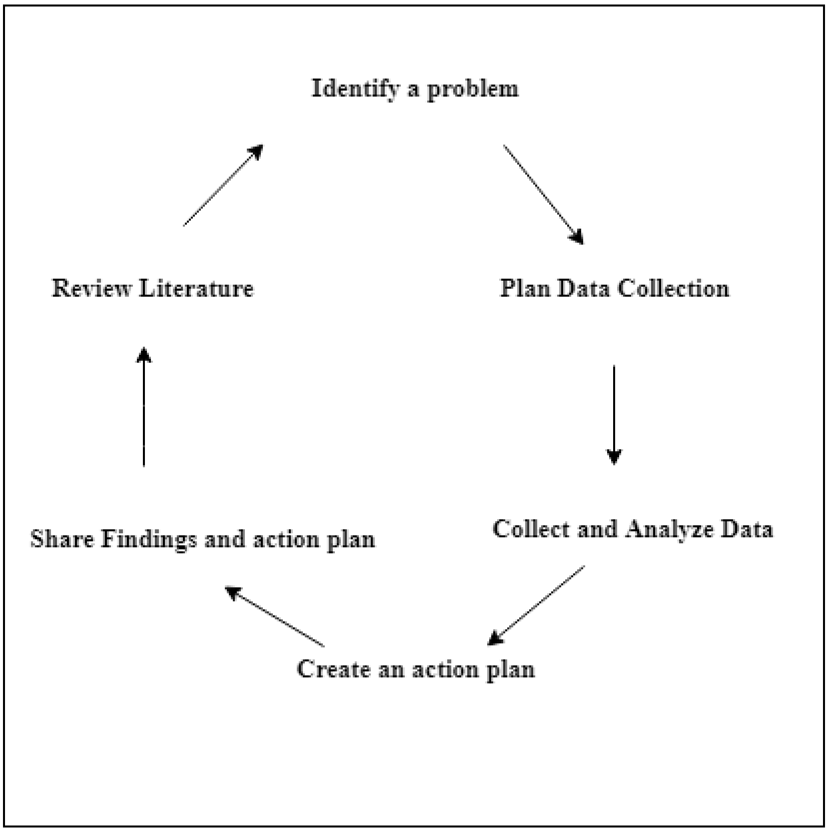

Figure 7 Steps for Action Research Process (Own image following (Johnson, 2012)*)*

3.2.2 Guidelines for performing Systematic Literature Reviews in Software Engineering (Kitchenham & Charters, 2007) and Procedures for Performing Systematic Reviews (Kitchenham, 2004)

What is Systematic Literature review?

According to (Kitchenham, 2004), a systematic literature review is a systematic way of identifying and evaluating the already performed research related to a specific topic or field of study.

A thorough and fair literature review performed systematically. It is considered to be of high scientific value, as the later research depends on the outcomes from this. It gives the basis for carrying out systematic reviews. Considerably taking more time and effort than a traditional review, the systematic review provides thorough information regarding the effects of certain phenomena's against research setting and empirical methods. If the outcome is consistent, this is the indication that the phenomenon is robust. In other cases, the reason and source of variation could be investigated.

A rigorously performed literature review in a systematic way would assist the research process in multiple ways

- Help to summarize and synthesize the existing research and evidence concerning to conduct and outcomes
- Identify research gaps identify and suggest the area for further investigation
- Helps to avoid duplication of work
- Assist in designing framework shape and the future research investigation plans

Review Process

A systematic review consists of multiple activities. The number and order of activities performed differ according to different guidelines. In the guidelines presented by (Kitchenham, 2004), the author has divided the review process in three-phase, with each phase consisting of multiple different activities.

1. Planning the review
2. Conducting the review
3. Reporting the review

1. Planning the review

The need for systematic review

Before beginning to perform a systematic literature review, the researcher must ensure whether it is needed. The need for systematic review arises when researchers want to summarize all the phenomenon related to the research topic in an unbiased way.

Specifying Research Questions

In the planning phase, the most crucial step is to define the research question. The researchers must have a clear and concise research question.

Structure of Research Question

Before performing SLR against the desired topic, the research question must be complete to follow **PICOC** method.

- **P**opulation, identified as to who? Population refers to the target group in the area of investigation
- **I**ntervention, defined as what or how? Intervention address the specific issue or area of research to perform an investigation
- **C**omparison, identified as compared to what? Comparison specifies what the investigation will be compared to.
- **O**utcomes identified as and specify what the required outcomes researcher want to achieve are
- **C**ontext represents the settings of circumstances and environment of investigation

Development of a Review Protocol

A review protocol states the underlying method which would be used to perform the systematic review. A structured and well-defined protocol will make sure the outcomes are not biased.

2. Conducting the Review

Identification of Research

The primary purpose of the systematic review is to find as many as possible related literature specific area of studies and research question. Generate search strategy to look for relevant literature in the database. Define keywords and generate queries to retrieve documents. Search strategies are mostly iterative.

Study Selection

After the relevant studies have been searched and collected, they will be assessed based on their relevance. A selection criterion is defined and the literature which meets that criteria are selected.

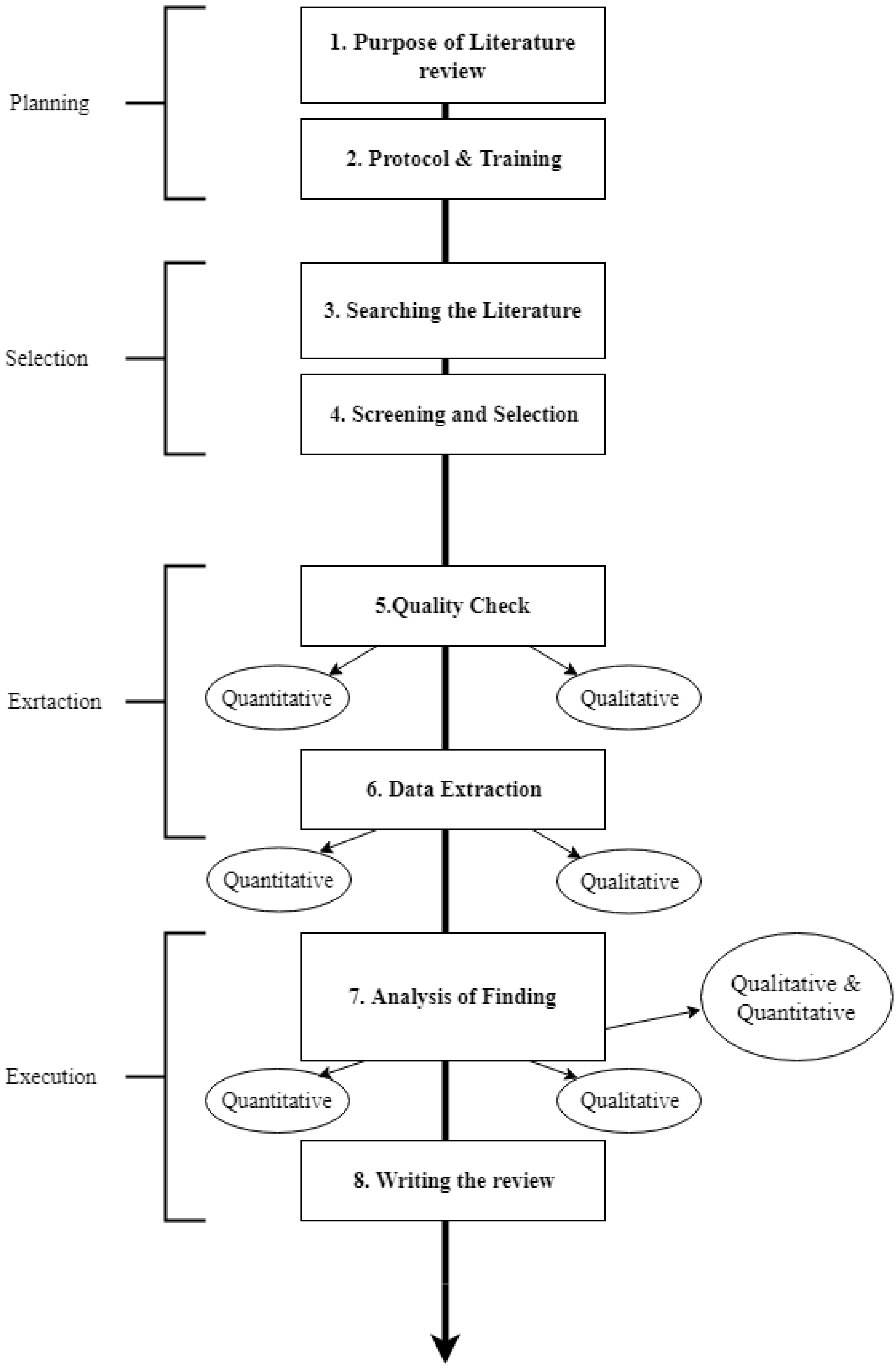

Figure 8 Steps to perform Systematic Literature Review (Own image following (Okoli & Schabram, 2010)*)*

Study Quality Assessment

After the literature is selected based on inclusion and exclusion criteria. The quality of the literature is assessed. The quality represents how much the research is bias and has maximum validation.

Data Extraction

Data extraction is performed by doing full-text studies of the selected literature and then record the information in the form.

Data Synthesis

The results from the primary studies are collected and summarized at this stage. Synthesis could be descriptive occasionally supplemented by quantitative summary. Using statistical techniques to achieve quantitative synthesis is called meta-analysis. The synthesis activities must be defined in the review protocol.

3. Reporting the Review

Effectively communicating the result from the review process is very important. To make sure that the later researcher could evaluate the finding and process study performed. The final report should contain all the details.

3.2.3 Scoping studies: Towards a methodological framework (Arksey & O'Malley, 2005)

There are several terms which have been used to describe the scoping review in literature including scoping study, systematic scoping review, scoping project, scoping report, evidence mapping and many more. In the available literature, it has also been defined by many authors in different ways.

According to the definition discussed in the paper by (Arksey & O'Malley, 2005), the scoping review purpose is to quickly record the critical concepts from literature which supports the research area with the evidence sources available. When targeting a complex area which has not been rigorously reviewed, it could also be considered as a standalone project.

As proposed in definition, scoping review involves a complete review of the literature. The extent of detailed analysis of extracted literature always depends on the research question and study purpose. There could be multiple reasons to conduct a scoping study, the four most common of them discussed below.

1) It is performed to study the variety and nature of research activity. In this type of research, findings are not discussed in details.
2) Sometimes mapping is performed to assess the feasibility of conducting systematic review relevant to the research area.
3) To describes detailed findings in the area of research study, and provide a mechanism to summarize those findings which the policymakers and consumers can quickly utilize.
4) Scoping review identifies research gaps by summarising the literature.

According to the methodological framework developed and adopted by (Arksey & O'Malley, 2005), consists of several stages before giving a conclusion or findings. There are a total of five stages

Stage 1 Identifying the research question

Stage 2 Identifying relevant studies

Stage 3 Study selection

Stage 4 Charting the data

Stage 5 Organising, summarizing and reporting the results.

Comparison of Systematic Literature Review and Scoping Review

Label	**Description**	**Search**	**Appraisal**	**Synthesis**	**Analysis**
Scoping Review	Initial assessment of potentially available research literature according to the scope of the study. The goal of this review is to recognise the extent of possible research evidence.	Search depends on time and other constraints.	No formal quality check	Mostly in tabular form with some narrative comments.	Describe the quality as well as quantity of literature by critical features. Tries to specify a practical review.

Systematic Literature Review	In SLR, it tries to search, appraise and synthesize research evidence systematically by following the provided principles and guidelines.	Perform an exhaustive and comprehensive search.	The inclusion and exclusion criteria are determined through quality checks.	The synthesis is narrative with tabular form.	Describe what is already known, the recommendations, unknown phenomena's and uncertainty about findings, and future research recommendations

Table 5 Comparison of Systematic Reviews with Scoping Reviews (Own table following (Grant & Booth, 2009)

3.3 Chosen Methodology

Systematic reviews are a type of research synthesis which is conducted to categorise and retrieve findings which is relevant to a specific research topic. In the following thesis, Systematic literature review is used as a research methodology to answer the research question.

As discussed in detail above SLR is suitable in the following situations to

1. There is a need to find standardized international evidence
2. To Check the current practice, variations to that practices or to classify new practices of the research area
3. There is a need to search, recognize and report the areas which have the potential for future research
4. Need to Inspect the conflicting results
5. Produce the statement to use as decision-making output

According to the following indication, SLR is best suited to our requirements (Munn et al., 2018). The SLR identifies and interprets already available research to answer the research questions by summarizing the findings and perform synthesis on those findings (Kitchenham & Charters, 2007).

The main steps during the research performed namely planning, conducting and reporting the review by following the guidelines proposed by (Kitchenham & Charters, 2007)

4 Systematic Literature Review

4.1 Planning the Review

In planning, research questions are proposed, a search strategy is defined along with search terms and search query, inclusion and exclusion criteria are defined. Following steps are explained in detail

4.1.1 Review Objectives

With already growing usage of agile methodologies in software development, data-driven organization have also put a focus on employing agile methodologies for analytics application development. In analytics, software development is not the whole process but a part of the analytics process lifecycle. Timely information is more critical than working functionality. Employing traditional methodologies have not reaped the same result as agile in software development. With successfully implementing the agile methods in software development, these methodologies could also be advantageous for delivering successful analytics projects (Nagle & Sammon, 2013). There have been several contributions to the research related to agile adoption in analytics application, but there is still no agreement and unobstructed views on the research topic. The main objective for doing this research is to have an in-depth understanding of agile practices in the data-driven environment and the challenges faced during the agile adoption. The agile best practices, enablers and success factors will also be discussed.

4.1.2 Research Questions

RQ1: How have agile methodologies evolved with the emergence of data science?

RQ2: Which agile frameworks, methodologies and best practices are well known for data-intensive applications and how they could be applied?

RQ3: What are the enablers and success factors for Agile Analytics?

RQ4: What are the challenges organizations faced for adopting agile methodologies for analytics applications?

4.1.3 Search Strategy

After proposing the objective for underlying research study and research question, a search strategy is defined to retrieve and analyze the already available literature specific to the research area. It involves determining the search sources which are mostly electronic databases. The studies were searched and retrieved from specified databases. The references to studies are also analyzed to discover other but related literature. After that, inclusion and exclusion criteria's are applied to the documents which were retrieved through search from the electronic database and reference search.

4.1.4 Search Criteria

The criteria defined to retrieve the literature from an electronic database consists of two parts. The first part of the search string includes words related to agile and the second part related to data analytics.

Database Resources

Database	URL
ACM Digital library	https://dl.acm.org/
IEEE Xplore	https://ieeexplore.ieee.org
SpringerLink	https://link.springer.com/
ScienceDirect	https://www.sciencedirect.com/
Google Scholar	https://scholar.google.com/

Table 6 Database resources for literature

Example Database search query

Agile AND ("business intelligence" OR "data analytics" OR "advanced analytics" OR "big data analytics" OR "BI" OR "data science" OR "data-driven" OR "analytics")

Keeping this search string as an underlying theme. For each database, the search query is generated according to the specification provided with that database.

In a standalone scenario, the following query is also applied.

"Agile Analytics"

There are also some queries used in combination with individual keywords from agile and data analytics. In the actual search queries will be modified according to digital database search requirements.

Agile Approaches AND ("business intelligence" OR "data analytics" OR "advanced analytics" OR "big data analytics" OR "BI" OR "data science" OR "data-driven" OR "analytics")

Agile methodology AND ("business intelligence" OR "data analytics" OR "advanced analytics" OR "big data analytics" OR "BI" OR "data science" OR "data-driven" OR "analytics")

Agile project management AND ("business intelligence" OR "data analytics" OR "advanced analytics" OR "big data analytics" OR "BI" OR "data science" OR "data-driven" OR "analytics")

Inclusion and Exclusion

From all the literature retrieved through a search query, the most relevant research which satisfies the inclusion criteria is selected.

At this stage, inclusion and exclusion criteria's are applied to the selected literature iteratively.

Inclusion

- The study and the research performed are in English.
- The study is conducted between 2002 to 2019.
- The paper should describe the agile method and practices applied in any of development phase of an analytics application.
- The literature gives insight into the adoption of agile in a data-driven environment.
- The study must be related to the research topic.
- Full text of the selected literature must be available.

Exclusion

- Literature which is not directly focusing on the agile methodologies and practices studies that do not discuss agile methods in context to a data-driven environment.
- PowerPoint presentations or slides, keynotes or tutorials.
- Agile methods applied in manufacturing or other industrial fields which do not involve data-oriented or analytics application.
- Duplicate paper from different databases.
- Research literature other than the English language.

4.2 Conducting the Review

In this phase, the findings and documents retrieved from the search and retrieval process will be discussed in detail

4.2.1 Search the Literature

The main goal of a systematic literature review is to find the possible number of literature related to the research area. To fulfil this requirement database search is performed according to search strategy using keywords and search strings. At first, preliminary searches were performed to examine the initial search results and fine-tune the query accordingly. Top results were compared with the most relevant papers which were already downloaded for testing. Searches were performed using the most common query string consisting of a single word or sentence. The search results were highly dependent on the keywords or search string used and were easily influenced by the minor change. Search is also performed to filter based on the titles. Though results showed that search string consisting of multiple keywords has some more precision than single keywords.

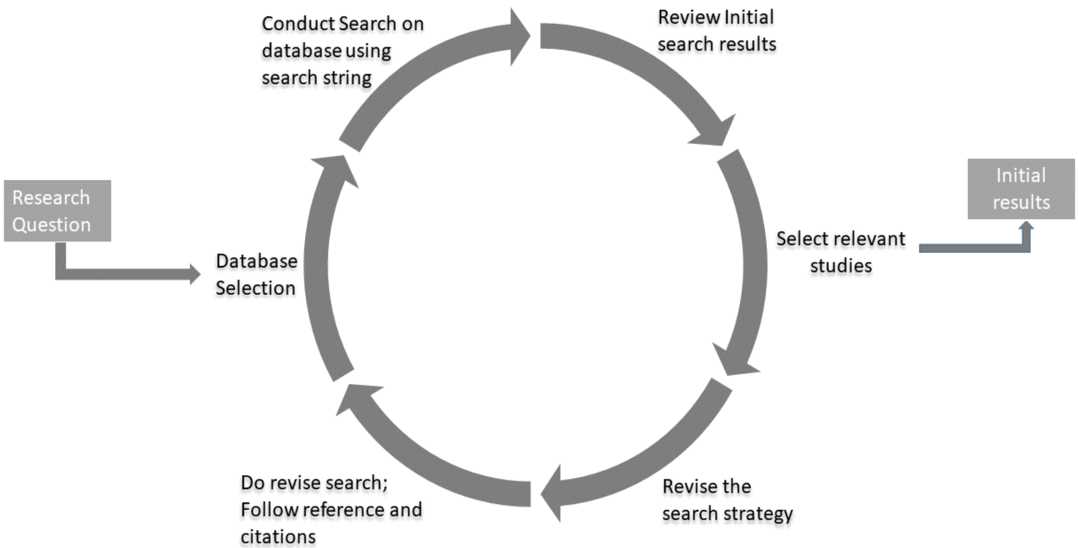

Figure 9 Iterative search cycle (Own image following (Jacobs, 2015)

After that, using the search strategy defined in the last step, the databases were searched, and related literature was retrieved. In the first step, there were vast numbers of results returned by each digital database, especially google scholar, which returned total in thousands. At google scholar, there were some duplicate references to other digital databases. From the original search from all databases, in total, a few hundred documents were retrieved.

4.2.2 Selection of Literature

After that inspection is performed based on inclusion and exclusion criteria. The selection and inspection process is iterative, in round one by extensive study of titles and abstract, documents which satisfy these criteria were selected. The search on all databases returned hundreds of papers in total. After this stage full-text filtering is performed. There were also vast numbers of irrelevant documents which were discarded. After the first round, we still had a considerable amount of papers left. We also have to apply exclusion criteria. After carefully applying the exclusion criteria, we left with a total 196 of documents. Doing the 3rd round of review, the final documents count is 47.

By referencing the base search strings and keywords as defined in the last steps, these strings were modified for each database. The search strings mentioned in Table 7 was not used precisely in the current form but modified and refined based on the search results. Like in SpringerLink, we explicitly filtered the result based on the word “agile” in the title. Similar modifications are also performed on other databases. Results from google scholar are not mentioned in the table as it returns a massive number of documents which are duplicate to already retrieved documents from different databases. In most cases, it is referring to the other database documents which we had already extracted. However, some papers which are retrieved from google scholar and added were checked through full-text search along with inclusion and exclusion criteria in the final round.

Search String	**ACM**		**IEEE**		**Springerlink**		**Science direct**	
	Total	***Selected***	***Total***	***Selected***	***Total***	***Selected***	***Total***	***Selected***
Agile AND (“business intelligence” OR “data analytics” OR “advanced analytics” OR “big data analytics” OR “BI” OR “data science” OR “data-driven” OR “analytics”)	12	**9**	690	**128**	95	**50**	96	**15**
"Agile Analytics"	5	**2**	4	**3**	14	**7**	9	**6**
agile AND ("data analytics")	5	**1**	61	**12**	1	**1**	686	**96**
agile AND ("business intelligence")	3	**2**	30	**9**	35	**17**	11	**4**
agile AND ("advanced analytics")	2	**2**	6	**2**	5	**1**	136	**32**

agile AND ("big data analytics")	*4*	**2**	21	**7**	31	**7**	2	**2**
agile AND ("data science")	*6*	***4***	*19*	***9***	*11*	***10***	*1*	***1***
agile AND ("data-driven")	22	**7**	34	**11**	40	**26**	*76*	**11**

Table 7 Query results after search and initial filters

Snowballing

Snowballing is a data-gathering methodology used in the SLR. The method utilizes the reference leads to get to more useful data. The reference list of literature is searched and followed to discover other useful but related sources (Wohlin, 2014). In the following research snowballing demonstrated good result to find content related to Agile Analytics. There was comparatively less literature retrieved through standard search because of Agile Analytics being a very new concept.

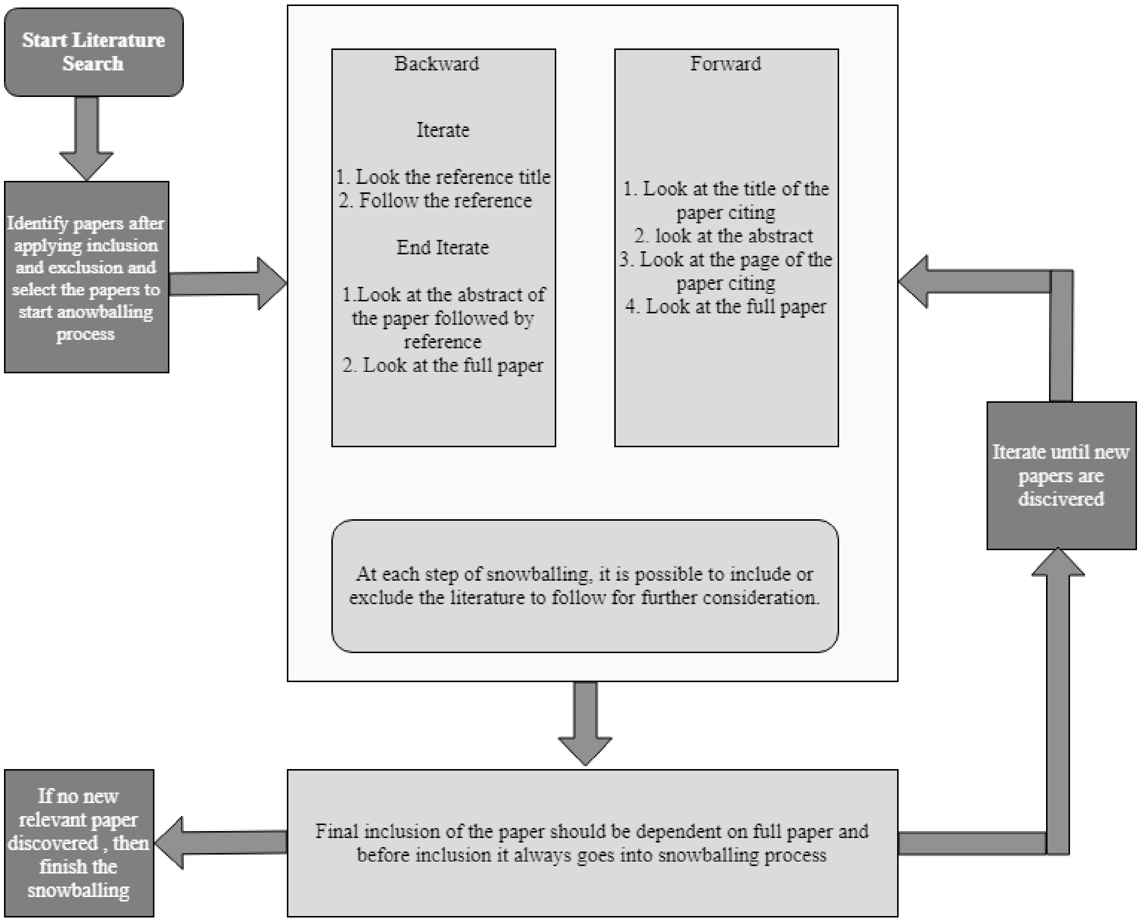

Figure 10 Snowballing step (Own image following (Wohlin, 2014))

Papers gathered after the search and applying filtration were evaluated, and references were followed. Forward snowballing was applied to identify referenced paper from already selected papers. This method of gathering the data is suitable for relatively new topics. Agile analytics is a new trend, and not much research was available on this topic. Also, in industry, it has not yet been implemented widely. With the help of snowballing, it was possible to get to the most relevant literature.

4.2.3 Literature Quality Assessment

Quality assessment criteria are applied to assess the methodological quality of the selected literature. It consists of a list of questions. The answers to these specific questions provide the measure to the extent to which the selected literature will contribute to the area of research.

The quality assessment following the criteria was applied for the reason that it will help to investigate the effectiveness of synthesis findings. This type of quality measures criteria's has been applied in systematic literature review since long (Dybå & Dingsøyr, 2008).

Each paper was evaluated according to quality assessment questions defined. Against every question, there is a rating scale. Scale range from 0 to 1 with research accurately answering the question rates 1, not answering at all rate 0 and partially answers were rated 0.5. The questions for quality assessments were

C1. Does the explicit context of the research defined in the paper?

C2. Do the objectives of selected literature are well defined?

C3. Does the research findings are ambiguous and are not clearly stated?

C4. How valuable is the area of study based on the research findings?

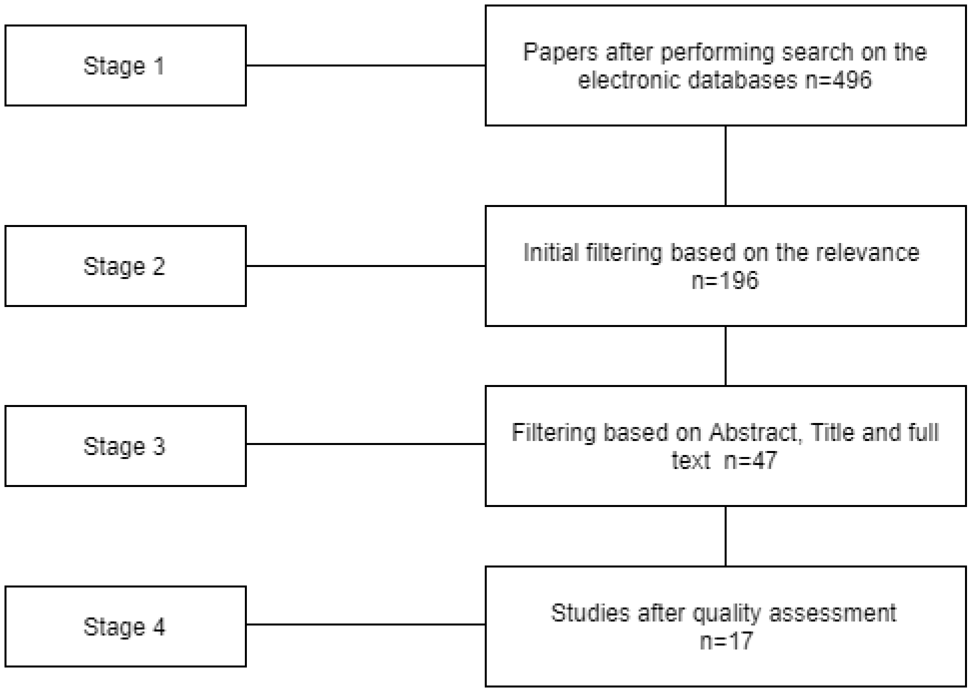

Figure 11 The selection and filtration stages of primary studies (Own image following (Ahmad et al., 2013)*)*

The quality assessment criteria were applied on all 47 papers, which were finally selected after applying inclusion and exclusion criteria and rounds of filtering. As the area of investigation related to agile analytics is entirely new, there are not many studies available. Therefore after applying inclusion and exclusion criteria's, moderate quality assessment was performed with a less strict criteria's. Criteria's are clearly defined and applied in context to the application of agile methodologies in the analytics environment. After performing the quality assessment, there was a total of 17 papers left.

Sr.No	Title of Article/Journal	Authors	Publication Year
P1	Agile big data analytics: AnalyticsOps for data science	Nancy W. Grady Jason A. Payne Huntley Parker	2017
P2	Comparing Data Science Project Management Methodologies via a Controlled Experiment	Jeffrey Saltz Ivan Shamshurin Kevin Crowston	2017
P3	Agile Project Management Approach and its Use in Big Data Management	Patrícia Franková Martina Drahošová Peter Balco	2016

P4	Agile big data analytics development: An architecture-centric approach	Hong Mei Chen, Rick Kazman, Serge Haziyev	2016
P5	A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science	Victor Chang, Deanne Larson	2016
P6	The Goal Questions Metrics for Agile Business Intelligence	Hadeer Sanaa, Walaa Afifi, Nagy Darwish	2016
P7	Big data analytics using agile model	Surend Raj Dharmapal, K. Thirunadana Sikamani	2016
P8	Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects	Kuldeep Deshpande, Bhimappa Desai	2015
P9	How to make business intelligence agile: The agile BI actions catalogue	Robert Krawatzeck, Barbara Dinter, Duc Anh Pham Thi	2015
P10	Agile Business Intelligence: Collection and Classification of Agile Business Intelligence Actions by Means of a Catalog and a Selection Guide	Robert Krawatzeck, Barbara Dinter	2015
P11	An Architecture for Agile Machine Learning in Real-Time Applications	Johann Schleier-Smith	2015
P12	Agile BI : The Future of BI	Mihaela MUNTEAN, Traian SURCEL	2013

P13	A Classification for Business Intelligence Agility Indicators	Henning Baars, Michael Zimmer	2013
P14	Business Intelligence and Analytics	Hsinchun Chen, Roger H. L.Chiang, Veda C Storey	2012
P15	Acceptance of agile methodologies: A critical review and conceptual framework	Frank K.Y Chan, James Y.L.Thong	2009
P16	Future research challenges in business agility - Time, control and information systems	Markus Strohmaier, Herwig Rollett	2005
P17	Beyond Flexible Information Systems : Why Business Agility Matters	Markus Strohmaier, Stefanie N Lindstaedt	2005

Table 8 Selected Journals and Conference Papers

4.2.4 Data Extraction and Synthesis

The data analysis is performed before extracting data. There are two methods of analysis discussed here.

1. **Literature Spread Sheet**

Referring to the guidelines proposed by (Kitchenham & Charters, 2007), after the selection of literature, the analysis was performed to identify and extract relevant information. In this process, an excel sheet was set up to record the core ideas and concepts, methodologies and findings the selected literature. Summarizing in the findings in the excel table provides a high level of understanding and clarification.

Following are the example data that could be extracted in addition to some other relevant details

- Title of the paper
- Authors
- Review date
- Relevance to the research area
- Methodology applied
- Data Analysis
- Validation techniques
- Limitations
- Future work
- Publication year

After the following data extracted from the papers, content analysis was performed to describe the nature of each study.

2. **Annotated Bibliography**

An annotated bibliographies are the list of summaries of the selected paper. The primary purpose of the annotated bibliography is to get a summary and evaluation of the article. The summary of the paper discussed the concepts, methodologies used, relevance to research, limitations and other important concepts related to our research topic.

Beside annotation, the bibliographic information of the article includes

- Article title

- Author name(s)
- Publication year
- Publication title (journal, report, book etc.)

There are multiple styles available to write bibliographic information. In this research, the APA style is adopted. *Annotated Bibliography* of all selected literature is performed and explained.

Chosen style

The *Literature Spread Sheet* style is a way of data analysis, which is more suitable when more than one researchers are working on the same topic, and there is a need to compare the results. In this case, thesis research is performed by one person. Hence ***Annotated Bibliography*** was adopted for data analysis.

Data Synthesis method

There are multiple methods for performing data synthesis, both qualitative and quantitative. In this research, the descriptive qualitative synthesis method is applied. Among several different methods available, two methods which are most suitable to this research type were discussed briefly. In the end, one method was chosen and applied to the collected data. The following two methods are discussed below

1. Meta-Ethnography
2. Grounded Theory

Meta-Ethnography

The primary purpose of SLR is to form the synthesis based on the findings. Synthesis is considered as a vital stage of SLR. Meta-Ethnography is an approach in which the main purpose is to perform synthesis by incorporating identified and selected concepts from multiple studies into the high-level theoretical structure. This approach is contrary to data aggregating quantitative approach like a meta-analysis. It has been applied vastly in health science, but there are very fewer approaches where it has been applied in software engineering or computing (B Silva et al., 2013). There are a total of seven steps of performing the synthesis of collected data

1. Getting started
2. Deciding the relevant literature
3. Reading the studies

4. Determining the relations between studies
5. Translating the studies into one another
6. Synthesising the translation
7. Expressing the synthesis

As there are multiple steps which have already been performed during the standard SLR process. The synthesis after the analysis is performed following these iterative steps.

Grounded Theory

Grounded Theory is a method of synthesis. The primary assumptions and features of the grounded theory include (Barnett-Page & Thomas, n.d.)

- It provides concurrent phases of data collection along with data analysis
- It is an inductive analysis approach
- The theory is developed based on the data
- The constant comparison method is utilized
- Use of theoretical sampling
- Generate the new theory

It is known as one of the best methods to produce theories. Grounded theory is well known for its ability to discover theory about the phenomenon with less explanation available about that phenomena. Grounded theory is sometimes also defined as the systematic way of discovering the theory from obtained data. It is ideally useful when to uncover social processes. The researchers employed grounded theory to find patterns and interactions among different social units (Chun Tie, Birks, & Francis, 2019).

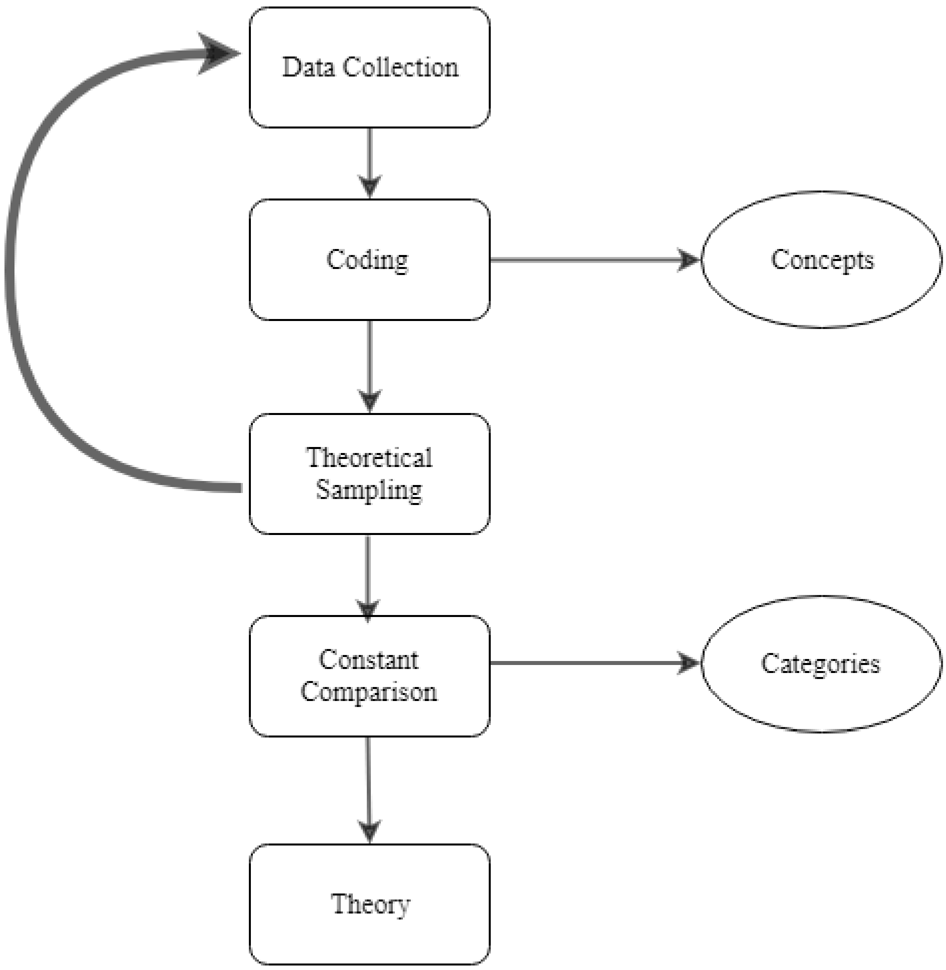

Figure 12 Grounded Theory Process (Own image following(Chun Tie et al., 2019)*)*

Characteristics	**Grounded Theory**	**Meta-Ethnography**
Focus	Developing the theory based on data	Describe a culture-sharing group
Best suited for	Best suited for understanding the interaction in social phenomena	Described shared pattern of a social group
Discipline Background	Sociology	Anthropology and sociology
Unit of Analysis	Process action and interaction of individuals	Culture sharing groups
Data Collection Forms	Interviews with individuals In the qualitative literature review	Observation, documents and artefacts

Data Analysis Strategies	Using coding	Theme and description of the case and cross-case studies
Witten report	Generate theory	The detailed analysis consists of cases

Table 9 Grounded theory and Meta-Ethnography difference (Own table following (Khan, 2014)

Chosen Method

By analysing the characteristic of both methodologies and suitability for the following research topic. The **Grounded Theory** method was chosen for data synthesis.

Concept Matrix

Concept Matrix is an effective technique to organize the research after writing annotated bibliographies and summarizing the concepts of all the papers. It is a technique employed by the researchers to present the association between the available research articles and the research topic. Each column in the matrix represents the concept intersection between different articles for a broader topic(Ochoa, n.d.).

Sources	Concepts			
	Concept 1	Concept 2	Concept 3	Concept 4
Source 1	✕		✕	✕
Source 2		✕	✕	✕
Source 3	✕	✕		
…. …		✕	✕	

Table 10 Concept Matrix (Own image following (Müller-Bloch & Kranz, 2015)*)*

Concept matrix supports in recognizing opportunities for synthesis. It describes the precise depiction of the research topic by analysing the overlapping concepts. Synthesis of all the article occurs when the bigger picture is displayed after answering the related questions.

- What are the aspects the authors agreed related to the research topic?
- What are the disagreements?
- What are the concepts which are discussed in specific articles but not in others?
- How has the research aspect been changed between different articles?

- What are the questions raised between articles?

By answer the following questions and some others, concepts are organized for synthesis purpose. Construction of the matrix also depends on personal creativity, originality and proficiency of the researcher (Klopper, Lubbe, & Rugbeer, 2007). Before the synthesis process, the concepts must be grouped and presented logically. Concept matrix is considered as an effective way of communicating findings and insights.

Concepts / **Title**	Agile	Agile BI / Analytics	Big Data /Data warehouse	Agile BI / Analytics Frameworks	Agile Analytics	Agile Methodology	Challenges of Agile BI/Analytics	Limitations of Agile BI/ Analytics	Agile Analytics Enablers or Success Factor	Practices	Difference to other agile methods
Agile big data analytics: AnalyticsOps for data science	×	×	×	×	×	×	×			×	
Comparing Data Science Project Management Methodologies via a Controlled Experiment	×		×			×				×	×
Agile Project Management Approach and its Use in Big Data Management	×		×	×		×		×	×	×	
Agile big data analytics development: An architecture-centric approach	×	×	×	×		×	×	×	×	×	
A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science	×	×		×	×	×		×	×	×	

The Goal Questions Metrics for Agile Business Intelligence	×	×		×			×		×	×	
Big data analytics using agile model	×	×	×	×		×		×	×		
Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects	×		×	×		×		×	×		×
How to make business intelligence agile: The agile BI actions catalogue	×	×		×	×	×	×		×	×	
Agile Business Intelligence: Collection and Classification of Agile Business Intelligence Actions by Means of a Catalog and a Selection Guide	×	×		×		×	×	×	×		×
An Architecture for Agile Machine Learning in Real-Time Applications	×		×				×	×		×	×
Agile BI : The Future of BI	×	×	×	×	×		×	×	×		×
A Classification for Business Intelligence Agility Indicators	×	×		×		×	×		×	×	×
Business Intelligence and Analytics	×	×	×	×		×	×	×	×		×
Acceptance of agile methodologies: A critical review and conceptual framework	×						×	×	×	×	×
Future research challenges in business agility - Time,	×		×			×			×	×	

control and information systems											
Beyond Flexible Information Systems : Why Business Agility Matters	×		×						×	×	×

Table 11 Concept Matrix

4.2.5 Present the Review

In the last phase, results from the literature review analysis are presented. The papers which were searched and collected from multiple electronic databases will be used to present the findings. The data was collected from the articles and papers using reading all text and writing short Annotated Bibliographies and later was synthesised by the data synthesis approach. After that, based on the finding, results are presented.

From 17 selected papers related to different aspects of the area of research, each paper was analyzed based on context, research question and empirical findings. The studies were conducted using multiple different methods.

Year-wise Selected Papers

From selected papers, the bar chart is displayed with the number of paper selected yearly in Figure 13. The image shows the number of papers published in a year from 2005-2017. Most of the papers, a total of 6 are from the year 2016. As the area of study is quite new, much of the work has been performed in recent years.

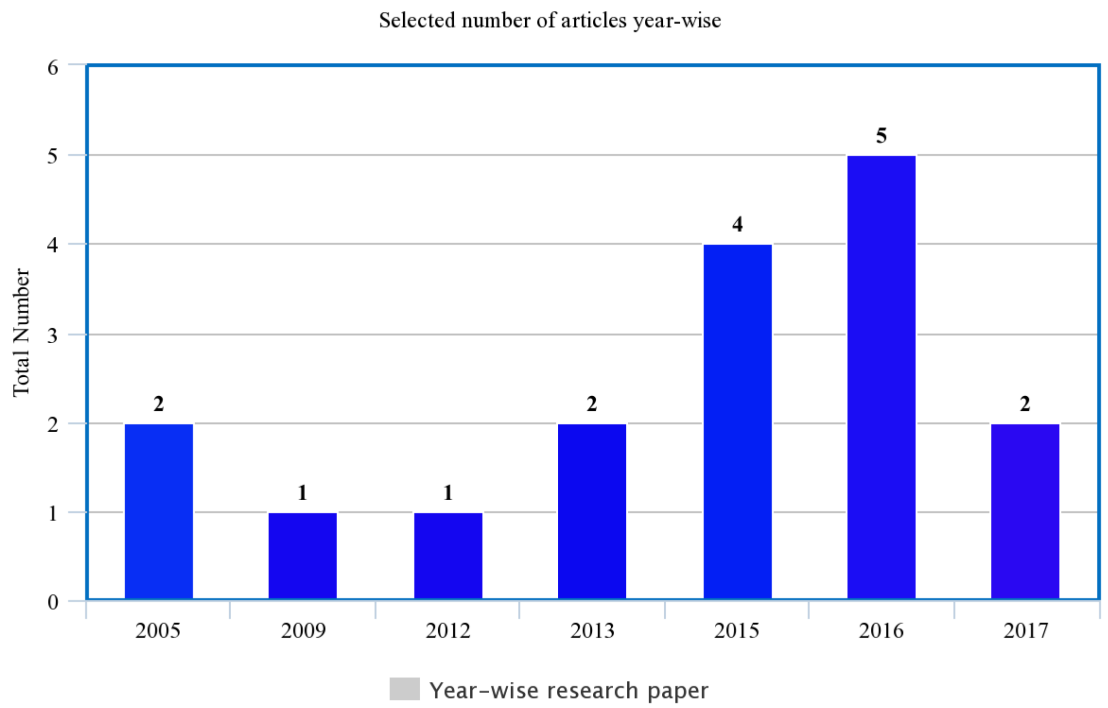

Figure 13 Year-wise selected publications (Own Image)

Research Methodology

There were multiple methodologies used in different research papers. Most of the selected paper has a different type of literature review as a research methodology. Total 58.8% of content have utilized this methodology. Detail of each research paper against the article is displayed below in the table

Case Study	Number of papers	Percentage
Experiment	3	17.6 %
Survey	2	11.8 %
Case study	1	5.9 %
Literature Review	10	58.8 %
Others	1	5.9 %

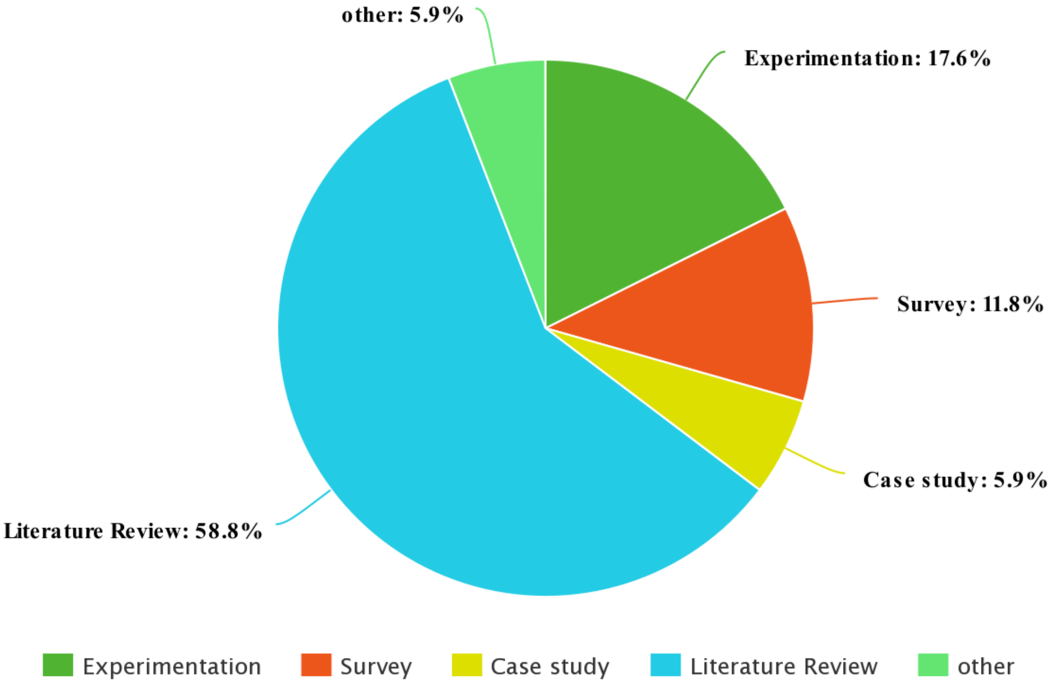

Figure 14 Percentage of methodologies applied in selected papers (Own Image)

Type of Research Articles

During the search and selection process, a considerable number of articles were selected and sorted, and finally, 17 articles were selected. Out of the following selected papers, total 8 were Journal papers and 9 were conference papers.

Type of Article	Total number	Percentage
Journal	8	47.1 %
Conference	9	52.9 %

Table 12 Type of Research papers

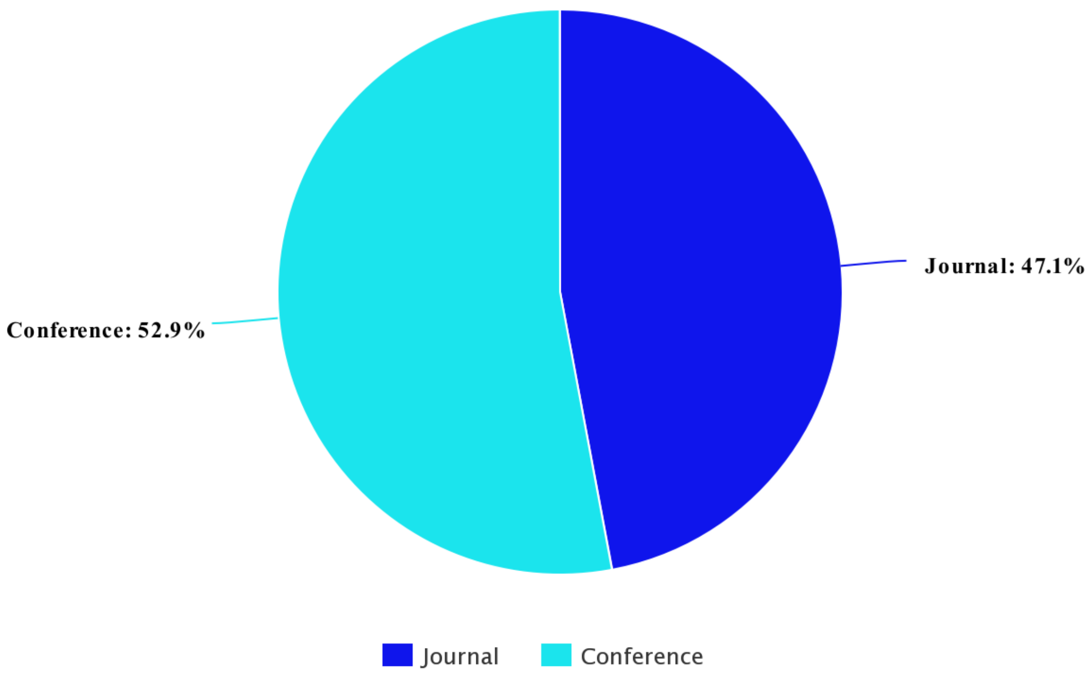

Figure 15 Research papers type (Own Image)

After collecting and summarizing all the papers. The following findings are extracted from each paper

Paper Title	**Summary/Excerpts/Findings**	**Research Approach**	**Type of Article**
Agile big data analytics: AnalyticsOps for data science	The author discussed the challenges faced during big data analytics lifecycle, especially with the introduction of advanced analytics. Traditional project management techniques cannot be applied in the same fashion. Machine learning and data science concepts need to be adopted with an agile approach to an analytics solution. The tailored agile approach should be adopted specifically to the project. The author proposed a new Data Analytics process model called Data Science Edge, which can adapt to new tools and technologies. It serves as a complete process model for knowledge discovery. By adopting agile models, the outcome would be quicker. The DSE proposed a conceptual framework which adapts to changes required in advanced analytics.	Literature Review	Conference

Comparing Data Science Project Management Methodologies via a Controlled Experiment	The author has compared multiple agile methodologies to measure and compare their performance on data science projects. There is not much information available to execute data science projects using agile methodologies. There is a need for well-defined process model specifically designed for data science projects. Different agile methodologies were tested for data science projects. According to experimentation result, Kanban turned out to be the most accepted methodology concerning factors like ease of use, project results, the satisfaction of individual team members and the willingness of the teams to work on future projects.	Experiment	Conference
Agile Project Management Approach and its Use in Big Data Management	The author compares different views from a management perspective to find the agile approach. The argument is about whether there must be a standard approach for all projects, or there must be a tailored approach for each different project. Big Data projects are the main focus point. The principles of agile manifesto are recommended for big data management. The author concluded that agile methodologies could be applied for smoother execution and delivery of big data projects. The agile approach is suitable for Big Data Management (BDM), considering the tools and technologies factors.	Survey	Journal
Agile big data analytics development: An architecture-centric approach	A methodology specific to Big Data Analytics is recommended. The methodology is known as Architecture-Centric Agile Big Data Analytics (AABA). Big data analytics is crucial to increase business efficiency. The author states the architectural challenges faced during the big data analytics process. Success is dependent on the performance and robustness of the underlying architecture. The other challenges are of adopting an agile methodology for the teams belongs to	Case Study	Journal

	different areas of expertise. Architecture centric approach by following the agile methodology is recommended for successful implementation. AABA methodology explained in detail. AABA being more architecture focused is slightly different from Agile Analytics.		
A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science	With the introduction of advanced analytics like data science, BI projects have changed the way of applying agile methodologies. Agile methodologies must be altered and tailored according to the changing requirements. The author summarized the changes required to incorporate new tools and technologies in the existing working environment. The author compares and proposed the new development lifecycle for advanced analytics projects. A framework for BI delivery is also presented by including the new advanced and fast analytics processes.	Literature Review	Journal
The Goal Questions Metrics for Agile Business Intelligence	This paper has discussed the agile business intelligence, agile BI framework and data warehouse. Explained the challenges and discussed the suitability of GQM for performance measurement of the agile BI teams. The GQM can guide the teams to achieve their desired goals. The questions are defined against the defined metrics. The answers to the questions provided the measurements against the team's performance.	Experiment	Journal
Big data analytics using agile model	The author provides guidelines about how agile methodologies could be applied to Big Data projects. The author explains each phase and step of analytics delivery and explains how agile methodologies could be adopted to each step. The author proposed three phases planning phase, development phase and closure phase. The author explains different roles in the agile analytics process.	Literature Review	Conference

Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects	The paper provides guidelines for assessing the suitable agile methodology according to the project's needs. Failure in data warehouse projects is due to using the traditional waterfall model. In some cases, agile methodologies have also not able to overcome failures, primarily because each project needs a tempered approach. Analytical Hierarchy Process (AHP) can be used to analyze factors which cause for selecting agile for data warehouse project. The author explained step by step approach to AHP. The steps include deciding for goals, comparison, comparison result and priority matric, accuracy check, and generalized model. The factors which cause selection of agile method are requirement status, organization culture and process, team effectiveness, data environment infrastructure.	Survey	Journal
How to make business intelligence agile: The agile BI actions catalogue	The paper discussed the competitive requirement in the business context, which is volatile and changing quickly. The author discussed how agile methodologies could be utilized to react quickly to the changing requirement in the BI system. A structured approach to react to change is missing in agile BI current approach. BI action catalogue is recommended, which consists of a set of action and steps needed to adopt agility in BI. The author divides the actions into four categories, which are Principles, process models, techniques, and technologies. Total of 21 actions is recommended.	Literature Review	Conference
Agile Business Intelligence: Collection and Classification of Agile Business Intelligence Actions by	List of actions is summarized based on the categories defined to increase the BI agility. Selection guide consists of a series of action that is recommended by the author. Schema is defined for different levels of selection guide. Thirty-one actions are recommended. Literature review	Literature Review	Journal

Means of a Catalog and a Selection Guide	methodology is applied to analyze and synthesize the collected data and summarize the results.		
An Architecture for Agile Machine Learning in Real-Time Applications	The use of agile for a machine learning project is entirely new. The author discussed the ways how machine learning for advanced analytics projects could embrace agile methodologies for smoother development and delivery. With the introduction of advanced techniques in analytics, there is a need for a new sophisticated process model which could be tailored according to project needs. Agile methodologies are applied in software development. The architecture, along with agile methodology approach, is recommended. The benefits of the recommended approach are improved collaboration, natural real-time processing and quick iterations. Concluding the article author asserts that agile methodologies could be used for implementing machine learning projects in data science teams.	Experiment	Conference
Agile BI : The Future of BI	The author tried to answer several questions related to agile BI. What is agile and how it is crucial for BI? What are the factors and essential features which support agile BI? The main components and layers of traditional BI architecture are discussed. The traditional approach is compared with an agile approach. Due to a massive set of problems with this approach, the agile approach is recommended with the benefits explained for each step. There are three main components of agile BI, which are agile development, agile business analytics, and agile information infrastructure. Agile BI is useful for changing user requirements, and to effectively handle the failures of IT system to meet the business needs, and fast delivery to market. The key elements to promote agile BI is recommended.	Literature Review	Conference

A Classification for Business Intelligence Agility Indicators	Meaning of agile is different for BI systems. BI functionality is a critical decision-support tool for top management and decision-makers. Agile is considered to be a practical methodology for BI development. There is not much work has been done until now on the following topic. The case study method is used with interviews from 14 experts. Industries with different areas are chosen for analysing how they are applying agile methodologies to their BI projects. BI architecture and agility of the target companies are studied. The agility of enterprise should not be measured with BI agility. BI agility must be measured after the requirements are precise and mature. BI agility is split between different organizational components. Agility must be divided and measured at each component. BI agility divided into three components BI functional agility, BI scale agility, and BI content agility.	Case Study	Conference
Business Intelligence and Analytics	The latest data technologies have increased the competition among the companies. Management needs to make quick decisions to be ahead in the competition. BI responsibilities have changed and evolved with data technologies. Emerging areas of BI&A are identified using the research framework. The BI&A is divided into three versions, BI&A 1.0 to 3.0. Each the aspects including application, data, analytics and impacts, were discussed. The current and future landscape of BI&A explained.	Literature Review	Journal
Acceptance of agile methodologies: A critical review and conceptual framework	A framework for applying agile methodologies have been recommended. With the adoption of agile methodologies in software, it has become smoother for teams to work iteratively in an incremental way. Much of the challenges handled efficiently with agile methodologies, which was	Literature Review	Journal

	not possible earlier with the traditional waterfall model. Earlier, organizations faced difficulty with adopting agile methodologies. The framework for acceptance of agile methodologies is discussed and defined. The factors involved in acceptance are motivation related factors like career consequences and organizational culture, training and external support, opportunity related factors, agile methodology characteristics, and knowledge management outcomes. Knowledge management could be used to examine the agile methodology acceptance.		
Future research challenges in business agility - Time, control and information systems	The author discussed the challenges faced by the organization in the form of competition. The businesses need to be able to adapt changes quickly. This ability to adapt changes quickly is called business agility. The dimension of problems concerning to business agility is time, control, and information system. There are two directions of business agility research, which are, organic IS and decision support system. The organic IS serves as autonomous IS. Decision support systems strengthen the human qualities in the control system. The triadic problem of time, control and information system will remain a challenge.	Literature Review	Conference
Beyond Flexible Information Systems: Why Business Agility Matters	The current state of the information system can handle the dynamic needs of businesses? Agility for any organization depends on how flexible are the information systems of that organization. Guidelines are proposed, which helps the businesses to adopt maximum agility. B-KIDE framework for business oriented modelling of organizational knowledge. The following framework address the knowledge to better understand process flexibility.	Literature Review	Conference

Table 13 Papers findings and summary

5 Conclusion, Limitations and Future work

5.1 Conclusion

From all the studies performed using systematic literature review on the application of agile methodologies in BI, analytics and DW, there exist multiple challenges in adoption due to less practical information available. The impact of Agile BI on business agility is not yet studied rigorously. The thesis aims to identify agile methodologies practices in Business Analytics environment, and the issues organizations faced and the new challenges with the introduction of new techniques and technologies. There has been much technological advancement in data storage and handling with data moving from relational databases to NoSQL databases. The new techniques and approaches like data science and machine learning have been introduced in analytics. There are multiple new and robust technologies for Big Data management. The agile methods have applied to most of the cases, and the outcome was better than the traditional approach. With customer involvement, every delivery after each iteration is a working functionality. The agile approach has addressed the problems of each stage of analytics lifecycle from data acquisitions to tools election, modelling issues and presentation.

RQ1: How have agile methodologies evolved with the emergence of data science?

Due to the successful implementation of agile methodologies in software development, the analytics and data science practitioners are trying to incorporate agile methodologies following agile manifesto. There have been multiple approaches to data science from a software development approach to architectural approach, and working on the development of a statistical model. All of these approaches somehow needed an agile approach with incremental iterations.

As compared to software development, data science requires more responsiveness. Agile promises this through the responsive development lifecycle and continuous feedback through validation. It makes sure that there is a continuous correction and also provides learning to the model. Teams collaborate to improve model performance. The agile approach provides all these characteristics by keeping the project quality and deadlines.

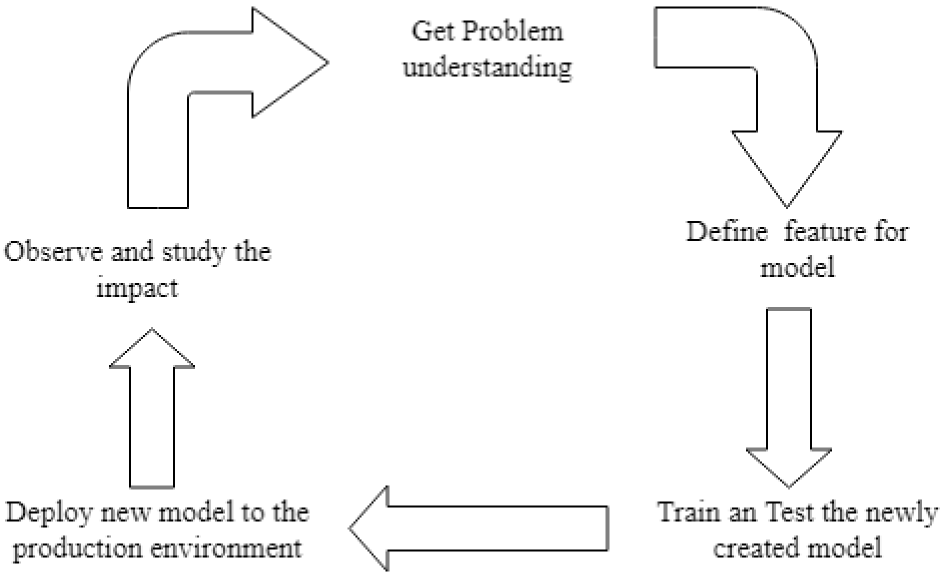

Figure 16 Agile Data Science lifecycle (Own image following (Schleier-Smith, 2015)*)*

As shown in Figure 16. The agile data science lifecycle starts with an understanding of the problem. The solution to the problem is proposed by creating new features. Training and test data is created and tested. After that model is deployed to the production environment, its performance and results are observed. If there is a need for improvement, a new iteration is started with new ideas and required improvements. The iteration will continue until the model is stable.

According to (Schleier-Smith, 2015) theoretical framework to perform data science with agile involves the following areas

- Iterative development
- Intermediate testing
- Prototyping
- Data visualization and understanding
- Structuring the data by data-value pyramids
- Follow the critical path
- Document the whole process

The theoretical framework also highlights the importance of iterative implementation, test and training of algorithm until deployment and later observation. In each iteration, intermediate work is shared with the teams to incorporate the feedback. In Figure 17, as explained the process path of retrieving the value from the data.

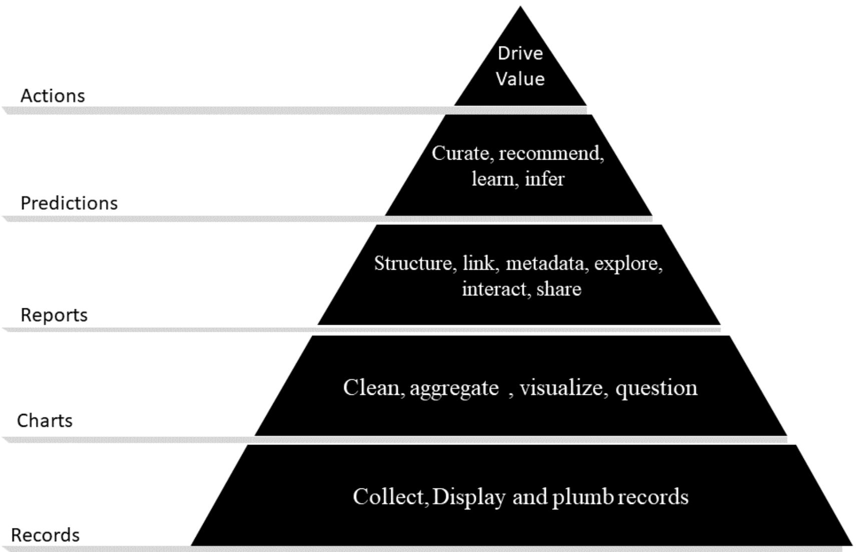

Figure 17 Agile Analytics Pyramids (Own image following (Schleier-Smith, 2015)*)*

The Data Scientist mostly work in the upper layer of the pyramid. The tasks performed at each layer are mentioned in the pyramid with the respective layer. At the prediction layer algorithms are trained, and in the end at action layer, discovered information is utilized to create business value (Schleier-Smith, 2015).

The process model based on CRISP-DM called Data Science Edge. The DSE recommends the BDA lifecycle and explains how it is different from BI. These are

- Scope
- Data Acquisition/Discover
- Analyze/Visualize
- Validate
- Deployment

Based on the experiments performed on the actual data science teams working in an agile environment at Walmart, a checklist is recommended consist of actions that should be performed before putting the code in production.

- The data scientist should be able to describe the accuracy of model and CPU consumption
- All type of debts must be minimized
- Everything should be documented
- The weakness of the model must be known to improve in the next iteration

DataOps techniques with tools and technologies have also been employed by data science agile teams to automate the processes and reduce the time of analytics development lifecycle.

RQ2: Which agile frameworks, methodologies and best practices are well known for data-intensive applications and how they could be applied?

There are multiple methods and frameworks available. The selection of the specific method or framework depends on the requirements. Irrespective of methodology or frameworks, teams in agile analytics use just in time planning with iterative and incremental delivery to incorporate sudden changes when the requirements change. After each iteration, the product evolves due to continuous involvement and feedback of the customer. Teams collaborate with the customer to towards required results.

Depending on the nature of requirements, BI type and context, there are multiple agile frameworks. To some extent, all of the frameworks and methodologies required tailoring according to business needs and type of BI system. Some of them are discussed below

Scrum

Used as a project management framework, Scrum is famous for its incremental nature of product delivery.

Kanban

It is a project management methodology used as continuous workflow management. Team’s activities can be visualized.

Test-Driven Development

In this approach, the developers write the test case and then develop the components to pass that test case. It ensures quality control in each life cycle.

Extreme programming

It is an agile framework which provides sustainable delivery within the teams.

As explained in the paper (Deshpande & Desai, 2015), the author explained the model to assess the suitability of applying agile methodology to the analytics projects. There are numerous aspects which need to be taken into considerations before moving towards agile.

In the paper (Sanaa et al., 2016), based on the best practices, the author recommended a framework for Agile BI. The framework consists of five phases.

- **Discovery** phase involves analysing the data sources and requirements
- **Design** phase involves designing the BI architecture
- **Development** phase objective is to deliver working software functionality
- **Test** included performing the final testing before deployment
- **Deploy** the working functionality after each iteration

DataOps or **AnalyticsOps** is the methodology recommended specifically for data-intensive projects which are using data science and machine learning. It enhances the fast time to value by promoting cross-functional collaboration. With an emphasis on people and processes, it enhances productivity by enabling agile and iterative workflow. The DataOps must be employed by several different functional groups to deliver business value. The following are the functional groups

- Data Engineers
- Data Scientists
- Developers and Architects
- Operations
- IT security and Data Governance

In another article, an architecture-centric approach called Architecture-Centric Agile Big Data Analytics (AABB) is recommended. It addresses both technical and organizational issues. It is different from other agile analytics approaches as it focuses more on architecture as a key enabler for success in agile adoption.

CRISP-DM and the extension called Data Science Edge (DSE) process model recommended by SAIC in the paper (Grady, Payne, & Parker, 2017). It serves as a model for knowledge discovery. As compared to processes of agile BI analytics. DSE has the following steps

- Plan
- Collect
- Curate
- Analyze
- Act

The planning phase involves the planning of analytics systems, data sources and requirements management. Collect phase handles the data collection process from different data sources. Curation of data is to map the data to integrate data into a central location. Analyze phase includes the design and development of the knowledge discovery. Last is the Act phase, which defines what actions have to perform based on the knowledge discovered?

RQ3: What are the enablers and success factors for Agile Analytics?

The key enabler for analytics projects of any organization depends on how the organization is planning to embrace agility. The business itself is agile or not. The data itself, selected methodology and the technologies have been the key enablers for agile analytics. The technology enabler has a specific focus on how the data is handled from data collection to data warehouse and its use for later in the machine learning or data science models. The availability of self-service capability and interactive visualization of analytics on a range of data. The users of the analytics projects are mostly management and decision-makers. They must be able to perform multiple interactive tasks without any assistance from IT teams. The technology advancement and available set of tools enable the user to retrieve the required result very quickly. There are multiple approaches to design among them we discussed is Architecture Centric Big Data Analytics (AABA) (H. M. Chen et al., 2016). Software architecture has a central role in adapting agile methodology. Although technology has a vital role in agile analytics, it heavily depends on organizational practices and how quickly and to what extent they are following agile. In this thesis and our findings, we have discussed multiple approaches of agile methodologies of adapting agile to analytics project. In other research questions, multiple methods and frameworks are also discussed and explained.

The success factors as explained in the literature, among them some we have discussed already in the literature are (Sanaa et al., 2016)

- Iterative development life cycle
- Flexible to Adopt Changes
- Value-Driven Development
- Production Quality
- Automate Routines
- Team Collaboration

- Self-Organization and Self-Managing Team
- Training and Coaching
- Customizing the agile approach
- Leadership
- Communication and transparency

RQ4: What are the challenges organizations faced for adopting agile methodologies for analytics applications?

There are multiple challenges involved in agile adoption. Some of them are specific to analytics projects, and some are related to organizations moving towards agile in generals. In case of analytics projects, if the organizations have not adopted to agile methodologies at all, then moving towards agile, they will face multiple challenges specific to analytics and agile adoption in general. Some of them are discussed below

User involvement

User experience is a critical component of BI and analytics. User must be involved in every phase of the implementation as possible to achieve the best results. All-time contact with the user is sometimes hard as the user could have a different schedule (Deshpande & Desai, 2015). To make the agile implementation plan which coincides with the user schedule is quite challenging.

Absence of Data ownership and Data governance

In most cases, organizations are struggling with information management. The identification of data assets and acquisition based on the ownership of data is quite challenging. For ongoing agile analytics process, a continuous availability, integrity, and security of data is of great concern and require many efforts.

There is no or minimal automation and TDD

Test-driven development is considered as a proven methodology for producing a quality product. Sometimes organizations have not adopted this process and also performing routine tasks manually rather than automating the process (Collier, 2011 p.225).

Lack of CI/CD practices

Like in software development DevOps technologies provide CI/CD capabilities which have fasten the time to delivery. CI/CD also improves collaboration among the teams, which is the fundamental principle of agile methodology. In analytics, adoption of CI/CD is relatively new and not much adopted.

The undecided priority of immediate or long term goals

When deciding to apply agile methodologies, Sometimes, organizations do not have a clear vision about the project and their organizational needs. They are always confused between immediate results or long term projects.

Technological difficulties in Agile Analytics

- **Tool support**
  There are multiple tools available for a different task in BI lifecycle. For TDD and ETL tasks, data warehouse and automation tasks, the available tools are not mature enough like in software development.
- **Data volume**
  The management of the volume of data is a great challenge. Providing the data for analysis from multiple data source is a processing challenge.
- **Heavy Lifting**
  There always needs much efforts in the backend to perform simple analytics tasks. Development of DW after collection of data multiple sources is a great challenge as it requires much efforts.

Organizational challenges

- **Lack of motivation**
  One of the organizational challenges is that the organizations are not motivated to adopt change. The benefits of adopting agile over current processes are not clear.
- **Complex hierarchy**
  A change in hierarchies and working environment needs a lot more efforts. Organizations sometimes stuck into a bureaucratic process and are not willing to move towards more flat hierarchies.
- **Not willing to invest**
  Moving towards agile needs massive investment in terms of hiring professionals, training and coaching and some other organizational stuff. The companies with a limited budget for an agile transformation are sometimes reluctant for such an investment.

5.2 Limitations

The idea of Agile Analytics is relatively new, and there is not much known about the impacts of applying it in the analytics projects. As there are not much actual case studies are available till now, this is considered as the limitation of the study as to know the real-time effects. However, the overall impacts which were proposed in the literature were studied. The organizational policies and the extent to which the organizations needs to tailor the agile methodologies are also not covered in this scope of the thesis. The other limitation was the absence of a quantitative approach. The literature sample set for a qualitative approach was also not diverse and did not have the detail about actual implementations.

5.3 Future Work

Agile Analytics is a relatively new context, and there have not been many publications on this topic. Application of agile in analytics has been there for a while, but most of them are tailored approaches of software development methodologies. As shown in our article search part, there has not been many results for Agile Analytics. The concept itself is entirely new and mixed with the induvial function or a part of the analytics life cycle. Much of the available literature was related to Agile BI. The use cases were from traditional Bi systems with obsolete requirements.

There has not been much discussion about new perspectives such as Data Science and Machine Learning in analytics, especially in BI. The nature and the context of the traditional agile approach are not producing the required results. Agile Analytics is the specialized agile approach to analytics, including BI and data warehouse projects. It has not been widely applied in the real world, and there has been less information available through practitioners. This topic needs extensive investigation as there are many areas which are still unexplored.

There have already been some attempts to explore the different context of agile analytics, including frameworks and methodologies, enablers and success factors, challenges organizations face during adoption, and the best practices involved in the adoption process. The research provided a good understanding of applying agile methodologies to Big Data Analytics. Agile methodologies are already making their space in the decision-support system, but there is still a need for improvement. In the following thesis, several different areas related to agile analytics are investigated and reviewed based on the available literature.

There could be many directions for further research, as the selected topic is very vast. The research in academia has started to rise, but there is not much theoretical work has been performed primarily in case of project management, there are very fewer case studies available. Still, there is a need to perform further studies through case studies as it provides real insight into the working and implementation of agile practices and provides with a point of view from industry practitioners. Studying on the organizational environment of companies where agile has been or going to be adopted in the analytics would give a whole lot of new information. The companies approach towards agile analytics along with technical capabilities could be studied. The outcomes could be compared with the ideal outcomes and investigate the challenges faced during adoption was tackled by respective organizations.

There could be further research on the methodologies, which has no proven results in real-time projects. The practices explained could be verified and further studies in different areas which are not studied in the literature. The behavioural barriers in the agile adoption within the analytics teams are also an area very less explored and could be a potential area for further research.

6 References

Abbott, L., & Grady, C. (2011). A Systematic Review of the Empirical Literature Evaluating IRBs: What We Know and What We Still Need to Learn. *Journal of Empirical Research on Human Research Ethics*.

Ahmad, M. O., Markkula, J., & Oivo, M. (2013). Kanban in software development: A systematic literature review. *Proceedings - 39th Euromicro Conference Series on Software Engineering and Advanced Applications, SEAA 2013*, 9–16.

Andersson, B., Beveridge, A., & Singh, K. (2015). *Literature Review, Academic Tip Sheet. 2003*, 1–2.

Angela Stringfellow. (2017). What Is Scrum? How It Works, Best Practices, and More - DZone Agile. Retrieved June 4, 2019, from https://dzone.com/articles/what-is-scrum-software-development-how-it-works-be

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*.

B Silva, F. Q., J O Cruz, S. S., Gouveia, T. B., Fernando Capretz, L., B, F. Q., J O, S. S., … B da Silva, F. Q. (2013). *Using Meta-Ethnography to Synthesize Research: A Worked Example of the Relations between Personality on Software Team Processes*.

Baars, H., & Zimmer, M. (2013). A Classification for Business Intelligence Agility Indicators. *Proceedings of the 21st European Conference on Information Systems (ECIS 2013)*, 1–12.

Barnett-Page, E., & Thomas, J. (n.d.). *Methods for the synthesis of qualitative research: a critical review ESRC National Centre for Research Methods NCRM Working Paper Series Number (01/09)*.

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., … Thomas, D. (2001). *Manifesto for Agile Software Development Twelve Principles of Agile Software*.

Cervone, H. F. (2011). Understanding agile project management methods using Scrum. *OCLC Systems and Services*, *27*(1), 18–22.

Chang, V., & Larson, D. (2016). A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science. *International Journal of Information Management*, *36*, 700–710.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88.

Chen, H., Chiang, R. H. L., Storey, V. C., & Robinson, J. M. (2012). *SPECIAL ISSUE: BUSINESS INTELLIGENCE RESEARCH BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT*.

Chen, H. M., Kazman, R., & Haziyev, S. (2016). Agile big data analytics development: An architecture-centric approach. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2016-March*, 5378–5387.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, *7*, 205031211882292.

Coelho Da Silva, T. L., Magalhães, R. P., Brilhante, I. R., De Macêdo, J. A. F., Araújo, D., Rego, P. A. L., & Neto, A. V. L. (2018). Big data analytics technologies and platforms: A brief review. *CEUR Workshop Proceedings*, *2170*, 25–32.

Collier, K. W. (2011). *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing: Delivering the Promise of Business Intelligence (Agile Software Development)*.

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In *New Horizons for a Data-Driven Economy* (pp. 29–37).

Deshpande, K., & Desai, B. (2015). Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects. *International Journal of Applied Information Systems*, *9*(8), 42–49.

Dharmapal, S. R., & Sikamani, K. T. (2016). Big data analytics using agile model. *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, 1088–1091.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*.

Earley, S. (2014). Agile analytics in the age of big data. *IT Professional*, *16*(4), 18–20.

Evelson, B. (2014). The Good The Bad And The Ugly Of Enterprise BI. Retrieved from https://go.forrester.com/blogs/14-08-25-the_good_the_bad_and_the_ugly_of_enterprise_bi/

Farhan, S., Tauseef, H., & Fahiem, M. A. (2009). Adding agility to architecture tradeoff analysis method for mapping on crystal. *2009 WRI World Congress on Software Engineering, WCSE 2009*, *4*, 121–125.

Firdaus, A., Ghani, I., & Yasin, N. I. M. (2013). Developing Secure Websites Using Feature Driven Development (FDD): A Case Study. *Journal of Clean Energy Technologies*, *1*(4), 322–326.

Grady, N. W., Payne, J. A., & Parker, H. (2017). Agile big data analytics: AnalyticsOps for data science. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-Janua*, 2331–2339.

Halper, F. (2017). *TDWI Self-Service Analytics Maturity Model Guide Interpreting Your Assessment Score 2017*.

Hossain, E., Ali Babar, M., & Paik, H. Y. (2009). Using scrum in global software development: A systematic literature review. *Proceedings - 2009 4th IEEE International Conference on Global Software Engineering, ICGSE 2009*, 175–184.

Hunt, J. (2006). Feature-Driven Development. In *Agile Software Construction* (pp. 161–182).

Ilieva, S., Ivanov, P., & Stefanova, E. (2004). Analyses of an agile methodology implementation. *Proceedings. 30th Euromicro Conference, 2004.*, 326–333.

Intelligence, W. B. (n.d.). *Where Agile Meets BI*.

Jacobs, S. (2015). Searching the literature is iterative. Retrieved from https://wp.nyu.edu/discoveringevidence/its-iterative/

Jasim Hadi, H., Hameed Shnain, A., Hadishaheed, S., & Haji Ahmad, A. (2015). Big Data and Five V'S Characteristics. *International Journal of Advances in Electronics and Computer Science,* (2), 2393–2835.

Johnson, A. (2012). A Short Guide to Action Research. *Special Education Department Faculty Publications*.

Kannan, K., & Services, I. P. (2011). Agile Methodology for Data Warehouse and Data Integration Projects. *Integration The Vlsi Journal*.

Khan, S. N. (2014). Qualitative Research Method: Grounded Theory. *International Journal of Business and Management*, *9*(11).

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.

Kitchenham, B., & Charters, S. (2007). *Source: "Guidelines for performing Systematic Literature Reviews in SE", Kitchenham et al Guidelines for performing Systematic Literature Reviews in Software Engineering*.

Klopper, R., Lubbe, S., & Rugbeer, H. (2007). The Matrix Method of Literature Review. In *Alternation* (Vol. 14).

Knabke, T., & Olbrich, S. (n.d.). *Towards agile BI: applying in-memory technology to data warehouse architectures*.

Ko, I. S., & Abdullaev, S. R. (2007). LNCS 4490 - A Study on the Aspects of Successful Business Intelligence System Development. In *LNCS* (Vol. 4490).

Krawatzeck, R., Dinter, B., & Thi, D. A. P. (2015). How to make business intelligence agile: The agile BI actions catalog. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2015-March*, 4762–4771.

Kruchten, P., & Gorans, P. (2014). *A Guide to Critical Success Factors in Agile Delivery*. 29.

Ktata, O., & Lévesque, G. (n.d.). *Agile development: Issues and avenues requiring a substantial enhancement of the business perspective in large projects*.

Labaree, R. V. (n.d.). *Research Guides: Organizing Your Social Sciences Research Paper: Writing a Research Proposal*.

Larman, C., & Basili, V. R. (2003). Iterative and incremental development: A brief history. *Computer*, *36*(6), 47–56.

Levy, Y., & Ellis, T. J. (2006). *Proceedings of the 2006 Informing Science and IT Education Joint Conference Towards a Framework of Literature Review Process in Support of Information Systems Research*.

Leybourn, E. (2013). *AGILE BUSINESS INTELLIGENCE OR HOW TO GIVE MANAGEMENT WHAT THEY NEED WHEN THEY NEED IT*.

Lim, E. P., & Chen, H. (2013). *Business Intelligence and Analytics : Research Directions*. *3*(4), 1–10.

Livermore, J. A. (2007). Factors that Impact Implementing an Agile Software Development Methodology. *Proceedings 2007 IEEE SoutheastCon*, 82–86.

Matharu, G. S., Mishra, A., Singh, H., & Upadhyay, P. (2015). Empirical Study of Agile Software Development Methodologies. *ACM SIGSOFT Software Engineering Notes*, *40*(1), 1–6.

Miller, H. G., & Mork, P. (2013). From data to decisions: A value chain for big data. *IT Professional*, *15*(1), 57–59.

Mohr, N., & Hürtgen, H. (2018). Achieving Business Impact with Big Data. *McKinsey*, 15.

Müller-Bloch, C., & Kranz, J. (2015). *A Framework for Rigorously Identifying Research Gaps in Qualitative Literature Reviews Completed Research Paper*.

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*.

MUNTEAN, M., & SURCEL, T. (2013). Agile BI The Future of BI. *Informatica Economica*, *17*(3/2013), 114–124.

Nagle, T., & Sammon, D. (2013). Fast and Flexible: Exploring Agile Data Analytics. *Cutter Benchmark Review*, *13*(2), 5.

Narasimhan, R. (2014). Big Data – A Brief Study. *International Journal of Scientific & Engineering Research*, *5*(9).

Newkirk, J. (2002). Introduction to agile processes and extreme programming. *Proceedings of the 24th International Conference on Software Engineering - ICSE '02*, 695.

Ochoa, A. (n.d.). *Concept Matrix for Sustainability Literature Review Aspect of Topic Natural Resource Levels Biodiversity Pollution Economic Impacts Urbanization How Topics Come Together*.

Okoli, C., & Schabram, K. (2010). *Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research A Guide to Conducting a Systematic Literature Review of Information Systems Research*.

Orton, C. G. (2015). Biological treatment planning. *AAPM 57th Meeting*, 1–5.

Paetsch, F., Maurer, F., Eberlein, A., Maurer, F., Eberlein, A., & Maurer, F. (2003). Requirements Engineering and Agile Software Development. *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE*, *2003-Janua*, 1–6.

Poppendieck, M., & Cusumano, M. A. (2012). Lean Software Development: A Tutorial. *IEEE Software*, *29*(5), 26–32.

Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*.

Sanaa, H., Afifi, W., & Darwish, N. (2016). The Goal Questions Metrics for Agile Business Intelligence. *Egyptian Computer Science Journal*, *40*(2), 24–42.

Schleier-Smith, J. (2015). *An Architecture for Agile Machine Learning in Real-Time*

Applications. 2059–2068.

Schniederjans, M., Schniederjans, D., & Starkey, C. (2014). *Business Analytics PRINCIPLES, CONCEPTS AND APPLICATIONS*.

SINGH, J., & SINGLA, V. (2015). Big Data: Tools and Technologies in Big Data. *International Journal of Computer Applications*, *112*(15), 975–8887.

Singh, S., Singh, P., Garg, R., & Mishra, P. K. (2015). *Big Data: Technologies, Trends and Applications*. *6*(5), 4633–4639.

Skene, A. (n.d.). Riting a Lit Review. *Writing*.

Souza, E., Moreira, A., Abrahão, S., Araújo, J., & Insfran, E. (2016). *Comparing Value-Driven Methods: an experiment design Adapt@Cloud: User-Centered Dynamic Adaptation of Cloud Services View project Requirements Engineering conference View project Comparing Value-Driven Methods: an experiment design*.

Stapleton, J. (1999). DSDM: Dynamic Systems Development Method. *Proceedings of the Technology of Object-Oriented Languages and Systems*, 406.

Strohmaier, M., & Lindstaedt, S. N. (2005). *Beyond Flexible Information Systems : Why Business Agility Matters*.

Sun, Z., Sun, L., & Strang, K. (2018). Big Data Analytics Services for Enhancing Business Intelligence. *Journal of Computer Information Systems*, *58*(2), 162–169.

Swapnil, W., & Anil, Y. (2016). Big Data: Characteristics, Challenges and Data Mining. In *International Journal of Computer Applications*.

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, *46*(6), 32–38.

The Data Value Chain. (2018).

Timulak, L. (2009). Meta-analysis of qualitative studies: A tool for reviewing qualitative research findings in psychotherapy. *Psychotherapy Research*.

Tom, B. (2012). Agile Analytics. *ACM SIGSOFT Software Engineering Notes*, *37*(6), 43.

Valentine, C., & Merchan, W. (2016). *DataOps: an Agile Methodology for Data-Driven Organizations*. 8.

Voigt, B. J. j. (2004). *Dynamic System Development Method*. (January), 22.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare the future. *MIS Quarterly*, *26*(2), xiii–xxiii.

Willans, B. (n.d.). *Writing a literature review*.

Williams, S., & Williams, N. (2007). *Critical Acclaim for The Profit Impact of Business Intelligence*.

Wohlin, C. (2014). *Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering*.

Woolley, B., Melbourne, V., & Hobbs, G. (2008). Agility in Information System. *Bseccanterburyacnz*.

Yin, R. K. (2015). *Yin, Robert K. - Case study research _ design and methods-Sage Publication (2015).*

7 Annotated Bibliography

Grady, N. W., Payne, J. A., & Parker, H. (2017). Agile big data analytics: AnalyticsOps for data science. ***Proceedings - 2017 IEEE International Conference on Big Data***

This article describes the challenges faced during the development lifecycle of an analytics application and with the introduction of advanced analytics. The author explains that the traditional way of handling the analytics cannot meet the changing environment and agile approach to decision-support information. How to incorporate Machine Learning and Data Science concepts in the analytics lifecycle and utilize agile methodologies in the information development lifecycle. The author explains how relying on the traditional data analytics lifecycle has created complexities and legacies and explains the needs to adapt to more agile and sophisticated processes. The author has adopted literature review methodology and industry practices. The author starts by explaining the limitations of the existing analytics process models. The author stresses the idea of "failing fast" and follow the direction which could promise the required outcome. The author explains how agile methodologies have been adopted for iterative and incremental software development and also describes the drivers causing that, among them are human factors, cost and risk factors and the system complexity. The author has proposed a new Big Data Analytics process model to incorporate new tools and technologies called Data Science Edge (DSE). It is extending and based on the CRISP-DM. It provides useful enhancement and serves as a complete lifecycle to knowledge discovery, which includes data storage, collection and software development. It provides the mechanism to decompose the steps to provide better enhancement and analytics improvement. The author has also discussed the theoretical frameworks to perform data science proposed by another researcher. The author compared the business intelligence and fast analytics with their proposed DSE process model. In the end, the author concludes that by following the agile methodologies, the outcomes could be expected more quickly to make a decision and to validate the current state of the project. The author explains how the DSE process model aligned itself similar to agile methodologies adaption in software development and proposed the conceptual changes required to adopt agile in data analytics.

Saltz, J., Shamshurin, I., & Crowston, K. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment. Proceedings of the 50th Hawaii International Conference on System Sciences (2017), 1013–1022.

In the following conference paper, the author has compared different agile methodologies to compare their performance concerning a data science project. The author has argued that with rapidly growing technologies and capability to process a vast amount of data, the data science field is proliferating. However, there is not much known about how to execute the project in a group. The author stresses the need for well-defined process methodologies for data science projects. The author has applied controlled experimentation methodology to test and compare the performance of different agile methodologies. The author chooses student groups as a subject and designed experimentation in a way that each group worked following different agile methodologies throughout the semester. The limitations are that the experimentations are not applied to real data science teams rather than students. In the end, the author analyzed and compared the performance results. The main research questions were to identify criteria to measure different methodologies and decide which methodology is better than others. In the start, the author did some literature review to build some theoretical background knowledge and explained the existing data science process methodologies like CRISP-DM and SEMMA and some other customized methodologies. The author explains with the help of a literature review that most of the data science projects are managed in an ad hoc way, with 55% of Big Data project never completed. The author compared the experimentation in the context of software development and after assigning the project and setting up the experimentation environment. The author did an interview and surveys and rated each methodology. The author provides pros and cons and also tried to extract the details about the environments where each method perform better than others. The survey results showed that the Kanban is the most accepted methodology concerning factors, ease of use, project results, the willingness of the teams to work on future projects and satisfaction of individual team members.

Chen, H. M., Kazman, R., & Haziyev, S. (2016). Agile big data analytics development: An architecture-centric approach. Proceedings of the Annual Hawaii International Conference on System Sciences, 2016-March, 5378–5387.

In this paper, the author has suggested a methodology targeting Big Data Analytics projects. The proposed methodology, called Architecture-centric Agile Big Data Analytics (AABA). The author states about the importance of big data for organizations to increase their operational efficiency, identify markets, making timely decisions and many more. The author highlights the challenges faced during Big Data implementation in an architectural point of view. The author explaining the technical challenges argued that success in any analytics system is directly dependent on robust infrastructure architecture that can effectively process big data. Organizational challenges include how the data scientist teams, along with software and data engineers, can work effectively to get the value from data. The author explained how agile methodologies could effectively tackle these challenges. In the paper, the author analyzed how big data system should be designed, which can adequately support the analytics process and how agile methods could be adapted for Big Data Analytics. The author proposed and tested the architecture-centric development, which, according to him, can support both, the big data system design and agile adaption. The author used an empirical case study method called (CPR) to identify practical issues and collaborate with the experts. The author analyzed 10 case studies from different organizations. The author explained AABA methodology in detail and described the software architecture as the key enabler for agility. The author explained the difference to AABA methodology to Agile Analytics, claiming it more architecture focussed. The author claims that this approach gives better control of continuous delivery. Although author claimed the success in the adaption of methodology, he clarifies that there are some biases and limitations including case study as a method itself, the use case studies from relatively big companies which are already in profits, outsourced part of projects and some more.

Chang, V., & Larson, D. (2016). A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science. International Journal of Information Management, 36, 700–710.

In this article, the author is investigating agile business intelligence and how the advancement in technologies and arrival of Advanced Analytics and Data Science has changed the way agile methodologies were applied in BI delivery. To adapt the new trend in BI agile methodologies must be altered to fit the changing circumstances. The author tried to understand and summarize how and how much agile practices have been transformed with new trends in BI and the challenges involved. The author starts by explaining the background information related to core topics and gradually discussing current agile practices in BI delivery. The author briefly compares the BI delivery and Fast Analytics and Data Science project lifecycle with detail to each step. After analysing both separately, the author described how fast analytics and data science could be adopted in Agile BI delivery framework. The author proposes two separate layers consists of a different task for Both BI and fast analytics and data science systems. The alignment between both layers with the impact the successful delivery of analytics system.

Sanaa, H., Afifi, W., & Darwish, N. (2016). The Goal Questions Metrics for Agile Business Intelligence. Egyptian Computer Science Journal, 40(2), 24–42.

In this paper, the author has argued the BI process and the challenges involved in that process. The paper aims to measure the performance of agile teams working towards the development of the BI system. The author has applied the Goal Question Agility Matrix (GQAM) method for this purpose. Developed by Victor Basilli in 1994, GQAM relies on BI and Agile concepts. The research methodology applied in this paper is experimentation as the performance of the teams will be measured against the set criteria defined by Goal Question Metrics (GQM). The author has explained about the essential components of BI systems, agile BI frameworks, agile BI success factors and tools and technologies required for success. The author explained all the phases of agile BI framework from the discovery phase to deploy. The author talked about the characteristics and attributes of the analytics application. In the end, the author applied GQAM and explained how it could be used to measure team performance. The GQAM is based on GQM, which has three components goal, question and metrics. It also provides the teams with guidance to achieve the desired performance.

Franková, P., Drahošová, M., & Balco, P. (2016). Agile Project Management Approach and its Use in Big Data Management. Procedia Computer Science, 83(Ant), 576–583.

In this paper, the author has argued the different views from the management perspective that whether there must be a tailored approach for every project or there must be the standard approach to all projects. The author put a particular focus on the management of Big Data projects. The author used the survey as a research approach. After performing interviews with management from different projects, the author identified and recommended the principles of the agile manifesto, which could be applied in big data management. Based on the research and data collected, the author tried to answer the questions about which approach is suitable for big data management, is agile approach is suitable for big data projects, and the recommendations for implementation of big data projects. The author starts from explaining the agile approach for software development and big data management process to make a theoretical background. Then the author compared the different aspects of BDM concerning their importance to the projects and concluded that agile methodologies could be applied to big data projects. By suggesting the agile methodologies for BDM author also advocated the other factors, including the technological considerations.

Dharmapal, S. R., & Sikamani, K. T. (2016). Big data analytics using agile model. International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016, 1088–1091.

In this conference paper, the author gave the background information about Big Data Analytics and discussed how agile methodologies could be applied to Big Data projects. The author argued about the importance of Big Data for business for competitive advantage. The author explains the companies can utilize the data for sales forecast, marketing and various purposes. The author describes the history and the rate at which the data is being created. The steps involved in the big data analytics from setting the goal to tool delivery are discussed. The author divided the whole analytics process into three phases planning phase, development phase, and closure phase. Besides, the author argued about different roles in agile development specific to big data analytics and compared the agile approach with the traditional waterfall approach. The author used a literature review methodology to extract findings and conclude results. The author concluded by explaining the adoption of agile in analytics projects. It provided collaboration, timely delivery to market and customer satisfaction.

Deshpande, K., & Desai, B. (2015). Model for Assessment of Agile Methodology for Implementing Data Warehouse Projects. International Journal of Applied Information Systems, 9(8), 42–49.

In this paper, the author has argued about the failure of data warehouse projects are due to its implementation using the traditional waterfall model. The failure ratio for data warehouse projects is very high. The author describes that although it is considered that agile methodologies can overcome this failure, but still, there are cases where agile methodologies have not made any difference. The author proposed that there must be an assessment for suitability of agile methodology before adopting it for the project. In this paper, the author has suggested the Analytical Hierarchy Process (AHP) to analyze the factors impacting the selection of agile for data warehouse projects. The author discusses the problems and challenges involved with the traditional waterfall approach to data warehouse implementation. Then the author explained about history and usage of agile methodology for data warehouse projects. By defining the objective of the study to identify the characteristics that impact the outcome of data warehouse project implemented using agile methodology, whether these characteristics affect positively or negatively, and what are the chances to predict the success based on that characteristics. The author then explained the steps for the AHP process, which include deciding for goals, comparison, comparison result and priority matric, accuracy check, and generalized model. The selection of BI projects suitable for agile methodologies depends on the four main factors, which are requirement status, organization culture and process, team effectiveness, data environment infrastructure. The author used the survey as a research methodology. In the end, the author proposed the model the assessment model and argued about applying it to as many projects to find the suitability score.

Krawatzeck, R., Dinter, B., & Thi, D. A. P. (2015). How to make business intelligence agile: The agile BI actions catalog. Proceedings of the Annual Hawaii International Conference on System Sciences, 2015-March, 4762–4771.

In this paper, the author is arguing about the changing requirement in the BI system, how the BI system would react in case of sudden and volatile requirements. The ability to react to change in BI systems is known as BI agility. The author conveyed that there is no structured approach to agile BI. To remain competitive for an organization, it must be able to respond to changing requirements. The paper aims to recommend the agile BI action catalogue. For any BI implementations, the practitioners can compare their actions with the catalogue to find missing steps. The author discussed the agile BI and divide the action required into four categories Principles, process models, techniques, and technologies. The author explained the step by step process of literature review and the analysis performed on the selected articles. For each category defined above, the author explains in details by collecting and summarizing the relevant literature to answer the questions. The total of 21 actions were recommended. The author argued that most of the actions identified are related to the implementation phase.

Schleier-Smith, J. (2015). An Architecture for Agile Machine Learning in Real-Time Applications. 2059–2068.

In this conference paper, the author put the focus on the more advanced techniques used by recommendation systems such as machine learning. With the advancement in the data handling tools and technologies and the emergence of data science and machine learning, there is a need to handle the project systematically as software development projects. Agile methods have been adopted for a long time and proved to be very beneficial in software development. In the case of machine learning agile approach is entirely new, and not much had been done till now. The author proposed an approach and architecture to handle the machine learning projects iteratively. Giving the example of some top tech companies like Facebook and hi5 explained their approach to machine learning and data science. The author summarized the approach and explained the key benefits, including improved collaboration, natural real-time processing and quick iterations. The author utilized experimentation as a research methodology and the Meet me application as a real-time implementation. By applying the event history architecture, the author concluded how agile methodologies could be applied in data science teams for implementing machine learning projects.

Krawatzecka, R., & Dintera, B. (2015). Agile Business Intelligence: Collection and Classification of Agile Business Intelligence Actions by Means of a Catalog and a Selection Guide. Information Systems Management, 32(3), 177–191

In this journal article, the author summarized a list of actions based on multiple categories to increase the BI agility. The author recommends the selection guide for BI actions required to adapt quickly changing requirements. The categories defined are principles, methods, techniques and technologies. The author explained the current background and importance of BI and explained the categories in detail.

Defined the schema for different levels of selection guide. The author proposed 31 BI action based on three levels and classification schema. The author applied the literature review methodology to analyze and synthesize the collected data and summarize the results in the form of selection guide.

MUNTEAN, M., & SURCEL, T. (2013). Agile BI.The Future of BI. Informatica Economica, 17(3/2013), 114–124.

In this paper, the author tried to answer the fundamental question related to BI development. The discussed the concepts which help to understand what is agile? Why is it appropriate to develop BI using the agile methodology? Also, which are the key features to support the agile BI? The author starts by explaining the basic concepts and tools and technologies for the BI system. The author defined BI according to different kinds of literature. The author explains the traditional BI architecture. The main components and layers in traditional BI systems include but not limited to, data sources, ETL layer, data warehouse, and BI tools for visualization. After a brief description of the architecture, the author explains its disadvantages. The disadvantages include a massive amount of duplicate data, use of different tools for every different task, rigid data models and more. The author then moves on to explaining the agile BI. The author breakdown the agile BI into three main components agile development, agile business analytics, agile information infrastructure. The author then explains each component in detail by comparing it with the traditional approach and performing a SWOT analysis for BA. After all the analysis performed, the author concluded that agile BI is useful for changing user requirements, to effectively handle the inability of IT to meet the business demands, and fast delivery to market. The author identified the key elements to promote agile BI.

Baars, H., & Zimmer, M. (2013). A Classification for Business Intelligence Agility Indicators. Proceedings of the 21st European Conference on Information Systems (ECIS 2013), 1–12.

In this paper, the author tries to distinguish how agile meaning is different for every BI solution. The author explains the BI function and its importance to top management. The author then continues by explaining what it means for BI to be agile and the effectiveness of using agile in BI systems. The author explained the related work from other publication and argued that on this topic, there is not much academic research has been done so far, and most of the information is available through practitioners. The author applied case study research methodology and interviewed 14 experts for data collection. In the case study method author choose the companies from different industrial background and which have already adapted. The author explored the architecture and agility applied to BI implementation of companies. The author concluded different result, the agility for the enterprise should not be tied to BI agility, BI agility should be measured after the requirement is matured and become concrete, BI agility should split into organizational components, BI agility should split according to each architectural layer. Overall, BI agility should be divided into three main components BI functional agility, BI scale agility, and BI content agility.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and analytics. MIS Quarterly, 36(4), 1165–1188.

In this paper, the author describes the importance of BI and relates it to the new emergence of the latest data technologies. The author proposed the framework to identify the emerging area of BI&A research. The author identified the critical challenges involved in this process. The author starts by explaining the key arguments and the facts related to BI, analytics and big data emergence and current level of use. The author has divided the overall BI&A structure into three versions, BI&A 1.0 to 3.0. The author then discussed the evolution and technology advancement in each version. There are new areas of research with the capability of new technologies to process and analyze the massive amount of data. The characteristics and capabilities of each version are discussed and summarized. The impact and usefulness of BI&A versions are discussed on areas like E-Commerce and market intelligence, E-Government and politics, science and technology, smart health and wellbeing, and security and public safety. Each aspect of BI&A like application, data, analytics and impacts were discussed. Current and future landscape of BI&A system are explained briefly.

Chan, F. K. Y., & Thong, J. Y. L. (2009). Acceptance of agile methodologies: A critical review and conceptual framework. Decision Support Systems, 46(4), 803–814.

In this paper, the author has performed a critical review of the framework for applying agile methodologies to a project. The author explains how the challenges can be handles while adopting agile methodologies for software development. The author used the literature review as a research methodology and provided a framework to perform future research. The author explains the introduction of agile methodologies in software development and explains the history of its arrival. The author discussed the problems organizations faced during adoption, which include developer ignore it, it was not correctly adjustable and more. The author explains the conceptual framework and the factors for acceptance of agile methodology which are ability related factors like training and external support, motivation related factors like career consequences and organizational culture, and the other factors are opportunity related factors, agile methodology characteristics, knowledge management outcomes. All these factors should be in consideration before performing the acceptance test of agile methodology. Knowledge management was proposed to examine the acceptance of agile methodology. The author explains how the characteristics explained in the conceptual framework can be used to understand the decision of developers to accept agile methodologies.

Strohmaier, M., & Rollett, H. (2005). Future research challenges in business agility - Time, control and information systems. Proceedings - Seventh IEEE International Conference on E-Commerce Technology Workshops, CEC 2005 Workshops, 2005, 109–115.

In this paper, the author discussed the challenges faced by businesses in the face of competitions for competitors. This paper conceptualizes business agility and identifies challenges for future research. The author explains the triadic dimension problems concerning to business agility that includes, time, control, and information system. Based on the following analysis, the author proposed two directions of business agility research, which are, organic IS and decision support system. The author starts by explaining the triadic problem in context to Business Agility and levels of analysis. The organic IS serves as autonomous IS, which is goal-oriented and decision support systems strengthen the human qualities in the control system. Due to the changing nature of Business agility the triadic problem of time, control and information system will remain a challenge.

Strohmaier, M., & Lindstaedt, S. N. (2005). Beyond Flexible Information Systems : Why Business Agility Matters.

In this paper, the author argues about whether the current state of the information system can handle the dynamic needs of businesses. The business agility for any organization depends on how flexible are the information systems of that organization. How quickly they respond to the change. The author performed analysis and proposed a guideline, which helps the businesses to adopt maximum agility. The author explains the business agility concept with the meaning of flexibility in business agility. The author applied B-KIDE framework for business oriented modelling of organizational knowledge. The following framework address the knowledge to better understand process flexibility. The author applied literature review methodology for analysis and proposed outcome.