International
Business Management
MEDWELL PUBLICATIONS SINCE 2000

A Study of Data Steward Involvement in Agile Big Data Project Management

Heba El Desouky, Sherif Mazen and Mohamed Elramly
Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

Key words: Agile, agile in big data, data steward, data governance project management

Corresponding Author:
Heba El Desouky
Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

Page No.: 247-255
Volume: 15, Issue 6, 2021
ISSN: 1993-5250
International Business Management

Abstract: The area of big data is attracting the attention of academia, industry and government across the world, due to the rapid development of the internet, the Internet of Things and cloud computing and the gigantic amounts of data collected daily. However, the question, "How to set up and optimize a big data project for successful assimilation of big data projects in organizations?" is still an open question. This is becoming an increasingly challenging task that requires understanding the value of big data, the challenges facing big data projects and the steps involved in running a big data project. Many researchers claim that failure rates for big data projects are considerably high due to several aspects affecting successful big data project management. In our study, we investigate the unique aspects of planning and managing big data projects. We propose an adaptation of traditional roles in agile projects to accommodate the specific needs of big data projects. Specifically, we propose and study the involvement of data steward role in big data projects to address the specific needs of such projects and eliminate the causes of failure, thus, increasing the successful completion rate of such projects. In this study. We back our position with a case study of the application of data steward in managing a big data project in a company working in the meal delivery industry. Our study lasted for a year and studied the performance of an agile project team represented by team velocity and productivity. Results show significant improvement.

INTRODUCTION

Data is becoming an important corporate weapon, businesses continue to expand their use and dependency on big data. However, success rates in developing and running big data projects are far less than other types of IT projects. In literature, various sources report that a minimum of 65% of Big Data Analytics projects fails. Reggio *et al*.[1] and Kabanda[2], reported similar findings to Gartner survey results[3]. The primary cause for the failure is not the technology but changing requirements, integrating with existing business processes and applications, management resistance, internal politics, lack of skills and security and governance policies[4]. Management of big data projects requires new ways of managerial thinking and new techniques of project management. Project managers are expected to embrace the digital transformation of their company and lead the transformation change[4, 5]. Kumar *et al*.[6] survey show that data analytics software products require more work because of their more complex nature. One of the possible reasons can be the volume and variety of data, the quality assurance for data analytics requires more time and a very different set of skills because software products used for data analytics are different from those used in traditional IT projects.

Data governance researchers proposed different ideas for helping big data projects overcome the challenges they face[7-9]. In the current world of enterprise-wide systems, business intelligence and metrics, as well as the security and privacy concerns that accompany them, are increasingly recognized as an institutional asset. This asset is of strategic importance and needs to be managed carefully and used appropriately. Data stewardship is providing a way to carry out this task in a coordinated fashion. Just as physical assets should be managed effectively, so, too should the data available to the organization about its services, programs, operations, finances and facilities. These need to be managed to improve understanding, increase efficiency and effectiveness, inform decisions and support change[10-12].

In our research, we study data steward role involvement in different stages of data projects managed by agile scrum methodology, through studying agile practices and philosophy for managing big data projects[13, 14] and the need for tailoring them to best fit big data projects[15-17]. We define this role in the context of agile-projects and verify its effectiveness through a case study.

Background

Big data: Big data analytics is all about tapping into diverse huge data sets and finding and monetizing unknown relationships. It is a completely data-driven process. Big data result from every business process involving a large volume of data that are available in many forms (variety). The speed (velocity), at which the terabytes of data (volume) are accessed, recorded, disseminated and used for further analysis is now becoming a big challenge for all decision-makers. Big data can be described as transactions, interactions and observations[18]. There are many new opportunities and challenges brought by big data. Some of these are not necessarily new but are issues that have not received the attention that they deserve[19].

Big data opportunities: How data-driven companies perform? Or, where's the evidence that using big data intelligently will improve business performance? Across all the analyses McAfee *et al*.[18] conducted, one relationship stood out: The more companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results. In particular, companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors. It was statistically significant and economically important and was reflected in measurable increases in stock market valuations.

The power of big data allows more accurate predictions, better decisions and precise interventions and can enable these things at a seemingly limitless scale[19].

Challenges in managing big data project: Big data projects are considered complex to be managed since they involve an unusual degree of uncertainty, unpredictability and produce frequent and small deliverables for inspection. The project's ultimate scope may be uncertain in the early stages of project development[5, 20, 21]. Today, digital data is often stored in many different formats, including unstructured databases and discrete text files[22]. Moreover, the volume of data is increasing day by day, making the task of handling data from various sources and in different formats even more challenging[23].

Managerial challenges of assimilating big data analytics are by various factors such as structural, technical, staffing and strategy-related issues which can be categorized according to the domains of technology, people and organization[24].

Tokuc *et al*.[5] investigated how the project management framework proposed by the Project Management Institute (PMI) can be effectively adapted to big data projects to reduce failure rates. Tokuc *et al*.[5], Zhou *et al*.[19] and Alharthi *et al*.[21], also researchers in[9, 25-28] addressed some challenges facing big data projects management including:

Multiple sources of data/Distributed existence: The difficulty of enforcing data consistency, the challenge of combining data from different sources arises from the different ways data is organized in various.

Multiple data formats: The non-clearly defined structure of much of the data. Data is frequently stored in many different formats, both structured and unstructured.

Privacy and rights management: Societal concerns over how data is being acquired and used are leading to increasingly tough regulations governing how organizations can acquire, store and use data.

Data trust: The huge amount, variety and velocity of data being created daily provides lots of opportunities and challenges for trust and reputation systems.

Data management/integration: With an average of 2.5 quintillion bytes of data created daily, businesses need to ensure it's going to the right people, applications and systems in real-time and that it's being used for accurate, transformative business intelligence.

Complex business relationship: A lack of data integration, an absence of data quality management policies and limited time and resources.

Agile processes: Agile processes represent a group of software engineering methodologies that promise to deliver increased productivity, quality and higher success rates in software development projects[29]. It is a software development philosophy and a mindset as well as a methodology which has been studied and researched for two decades[30]. The spirit of agile methodologies is delivering value to the customer frequently in small increments through short time-boxed development cycles. This is achieved via. a self-organizing team, high interaction with the client, welcome to change and adaptation and a set of disciplined practices to ensure product quality. Some different agile methodologies are available to choose from depending on the needs and requirements[31]. Currently, dominant agile methodologies are Scrum and Kanban.

Agile methodologies in big data: Many researchers suggested for agile practices and philosophy to be used for big data projects as they were successfully used in and transformed software development. They solve some issues inherent in the highly linear approach of waterfall methodologies. Because of the experimental nature of analytics development, detailed requirements cannot be set with complete confidence. It is only when the results are meeting the needs of the organization that the details of the end-state analytics models become clear[13, 14]. However, tailoring is required and new skills need to be introduced to agile practices to best fit big data projects[16, 17].

Data governance: A data steward is the operational aspect of data governance where the actual day to day activities of work gets done. Governance ensures that business needs are clearly defined, agreed upon and satisfied appropriately. Governance sets the priorities and how decisions are made; it monitors performance and compliance against the agreed objectives[26].

Data governance is not a technical application but about policies, organizations, standards and guidelines. The introduction of technology without responsible organization or policy preparation can increase risks. Data governance is needed to provide and share accurate and complete information about the current status with stakeholders[9]. Key motivations and early benefits of implementing data governance include[32]:

- Improved accountability to produce high quality and reliable data (sources of truth)
- Ensuring that the data are accessible and integrated using a common linear referencing system
- Engaging business areas within transportation agencies in their data, rather than viewing data as strictly an information technology issue

Data governance in big data: There must be clearly defined business objectives for big data use, the compliance objectives and the acceptable levels of risk must be set at the board level andthe responsibilities for big data must be clearly defined and it must be possible to measure how well the business objectives have been met[9, 26]. Data standardization, policies and processes and organizations are key components of data governance[33] with data attributes as ordinary dataaccessibility, availability, quality, consistency and security[34, 9].

Data that may have long-term value should be documented, referenced and indexed so that others can find and use them accurately and appropriately.

Literature review: This section reviews the few research works that discussed the need for data stewards in big data projects. Gharaibeh *et al*.[32] studied data governance within a sample of transportation agencies. They found that a pyramid-shaped data governance structure is commonly used. This structure consists of a lower level of stewards accountable for the quality and use of individual information technology. In their study, they pointed out the key motivation for introducing data steward role through interviews without taking into consideration the type of project, complexity and used practices, especially for complex projects like big data.

Plotkin[35] discussed in detail howto implement successful data stewardship in business and IT and how data stewardsare the backbone of a successful data governance implementation. He used metrics for measuring data steward progress. However, the book focuses more on roles, responsibilities and measuring impact on data quality and other aspects but didn't cover the methodology used, challenges faced and other project management implementation aspects.

Kim *et al*.[9], Small[26] and Gharaibeh[34] and other researchers focused on data governance and data steward aspects related to data quality level. They focus more on timely, reliable, meaningful and sufficient data services focusing on what data attributes should be achieved based on big data analytics. However, they didn't cover other challenges and daily activities and needs for the agile team.

MATERIALS AND METHODS

Data steward: The data steward role is still fairly new as a broad concept. The term is used inconsistently from one organization to another and in the literature.

A data steward is a role within an organization responsible for utilizing an organization's data governance processes to ensure the fitness of data elements: both the content and metadata[36]. It encompassesall activities that preserve and improve the information content, data quality, accessibility, ensuring compliance with data rules and regulations and usability of data and metadata[32, 36, 37]. Data stewards work with security, privacy and compliance officers to ensure that data are classified appropriately and that appropriate training is provided to users who will interact with the data[10].

According to Gharaibeh *et al*.[32], enterprise data stewards are a group of individuals who facilitate cross-subject area and cross-business unit priorities, projects and agreement and act as champions of data governance within their program areas.

Data steward's responsibilities: The responsibilities of a data steward may vary depending on the organization's size and type, as well as data maturity (e.g., whether data governance is in place or not). For example, the term data steward may refer to an IT staff member who is responsible for maintaining control of access to data, to a staff member who updates particular data elements, to a department director who manages her staff's use of the data, or even to a committee that approves new data values. Data steward's responsibilities can be grouped into four main areas[10].

Operational oversight: S/he oversees the life cycle of a particular set of institutional data to help in day-to-day decision-making.

Data quality: S/he is responsible for establishing data-quality metrics and requirements, including defining the values, ranges and parameters that are acceptable for each data element.

Privacy, security and risk management: S/he establishes guidelines and protocols that govern the proliferation of data.

Policies and procedures: S/he defines policies and procedures for data access, including the criteria for authorization based on the role and/or the individual.

In summary, effective data stewards make data available to the institution, thereby playing a key role at the heart of collaboration, supporting institutional research, assessment and analytics efforts that involve the domain data[10].

Data steward's skills: In addition to the professional traits that a good data steward should possess, generic and role-specific skills are needed. Some of these are common to all domains and some are specific to certain areas of expertise[10].

Required skills include facilitation, process definition, problem-solving, communication and relationship management. These will go a long way in making the steward as effective as possible. Given the proper skill set, a data steward can be a considerably valuable asset for the organization[10].

Contribution-data steward involvement in agile projects: Setting up and optimizing a big data project is challenging as explained in previous sections. There is a high potential for resolving part out of the mentionedchallenges by defining how the role of data steward can be part of mandatory skills to set up a big data project.

In Fig. 1, we show our proposed team structure by involving data steward in the data project addressing different challenges which is important to reach successful implementation for big data projects.

Furthermore, it's important to make sure that data steward, while embedded within the project, work together with the team defined in agile project management, all of them as team should share the same goals, creating a peer support network and being able to utilize complementary skills of the different team members.

Since, Scum is the currently most dominant agile process[38], we worked on integrating a data steward role into Scrum. Three roles are defined in the Scrum agile process, the self-organizing development team, the scrum master and the product owner. The product is built in a series of fixed-length (usually two weeks) sprints that deliver quick value to the client and are planned according to his priorities.

The scrum master is the facilitator and leader of the team. A product owner is responsible for defining the product requirements and their features and communicating with the client. However, data projects impose new challenges which require a need for more data related roles to solve these challenges. Product requirements are collected in a product backlog. For each sprint, a spring backlog is drawn from the product backlog depending on the priorities and goal of each sprint.

Four important meetings happen in Scrum. These are Sprint Planning meeting for the entire team to collaborate in defining sprint goals and priorities, Sprint Review meeting to review what was accomplished in the product and present it to the client, Sprint Retrospective meeting to discuss what went well and what went wrong in the sprint in terms of the process and finally, Daily Standup meeting to discuss how team members are progressing towards the sprint goals[38].

As a software development methodology, Scrum has nothing specific related to big data projects. No role or special events are related to data collection, quality, management or governance.

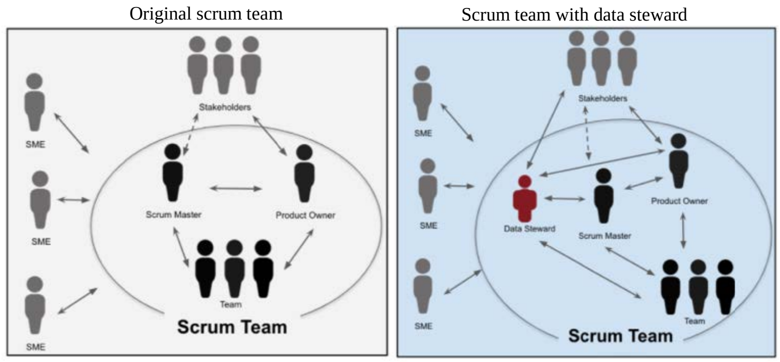

Fig. 1: Original vs. proposed scrum team

For the application of Scrum in managing big data projects, we study how can a data steward address these issues and carry out the responsibilities mentioned in the previous section. However, his/her responsibilities within a Scrum team may differ from one team to another. But in general, this role will include the following,

During the project inception phase, S/he is responsible for compiling a list of data sources to be used in the project and in developing the business use cases. S/he also would work on data quality, data-related requirements and privacy and security aspects of the project.

In the sprint planning meeting, when the iteration is planned, S/he provides input about data quality, facilitates data access procedures and carries out other activities of her/his responsibilities listed in the previous section.

During sprints, the data steward carries out her/his responsibilities defined before like defining the rules and policies for accessing data, developing privacy controls and ensuring their implementations, etc. Her/His specific tasks in each sprint will be based on the user stories chosen for this sprint.

In Sprint Review and Sprint Retrospective Meetings, the data steward contributes like other team members and gives his input.

Daily scrum meeting is not mandatory for data steward; however, s/he may join but is not expected to contribute except is asked to provide input or asking for help.

Facilitation and mediation: The data steward also serves as a mediator between the team and other parties in the organization involved with the data. This ensures that the development team focuses only on development related tasks and all data governance tasks are being handled by the data steward.

Data stewardship is the nexus of data governance. It provides linkage among owners of different but related data subjects. And it connects business rules and requirements with data models, database design, information systems implementation and day-to-day management and administration of data[39].

Case study: To enrich our study, we have conducted a year-long industrial case study. In this case study, we applied our proposal to a big data team within an organization struggling with data governance related topics, like many other organizations, we applied our proposal in setting up agile practices for their big data projects. The goal was to resolve part of challenges by defining how data steward can be involved during the project management. We have experimented on four cycles, studied and measured the effectiveness of data the steward involvement.

In first cycle, we built the baseline of learning more about the organization's structure, the culture that affects how decisions are made, data teams' day-to-day activities, applied agile methodologies and more information for building research baseline and helping to set up successful data steward role.We have explored different challenges they are facing with using agile scrum methodology for their project, we have used action research methodology[40] which is an iterative process involving researchers and practitioners acting together on a particular cycle of activities, including problemdiagnosis, active intervention and reflective learning. The cycle duration was three months each with spirals of planning, action, observation and reflection, the spirals are interwoven, fluid and repeated throughout the investigation.

An initial feedback workshop conducted where the researcher and data team members including the data steward collaborated together to learn more about data governance needs from their points of view which can solve part of pain they face. The workshop was with benefit for the researcher, the team and the organization itself. The feedback was commonplace in the workplace and we learnt more about real needs to ensure maximizing success ratio. The researcher was the meeting facilitator responsible to chair discussions and maximize participation.

Workshop setup started by presenting data governance elements and how it can help organizations to resolve their needs, followed by a collaborative workshop to hear and discuss more from the team their ideas for improvements and pain points which need to be resolved,

Table 1: Case study setup

Organization size	1300 employee
Industry	Meal delivery industry
Business domain	Traded meal-kit, operating in 11 different markets with one main technology office
Location	Berlin, Germany
Number of teams	~ 55-60
#Number of teams using agile	~42
Data teams	Data warehousing Data Analysts Data Infrastructure
Data teams using agile	Data warehousing Data infrastructure
Case study duration	From January 2018 to December 2018
Historical data available	From January 2017 to December 2017
Tracking system	JIRA

Table 2: Case study blocked work items

Month	Blocked	Resolved	Resolved vs. blocked (%)
Apr.	20	25	80
May	22	25	88
Jun	8	12	67
Jul.	7	16	44
Aug.	15	28	54
Sep.	20	29	69
Oct.	5	11	45
Nov.	5	10	50
Dec.	6	10	60
Average improvement			61.82

we have used the easy-to-use smart display tool called "Jamboard" during the workshop. The workshop built shared understanding about things that are important to the success of data foundation products and solutions being developed within the organization. And it cleared up the need and importance data governance can bring to the organization.During the workshop, we categorized the points into four main categories, Process related, Accessibility standards and processes, trust-ability (data quality and incidents) and find-ability and definition.

After the workshop, the researcher collaboratively with the data steward prepared data stewardship three quarters roadmap with the strategic plan that defines a goal or desired outcome and includes the major steps or milestones needed to reach it, using workshop outcome as input. Steps followed to create the roadmap was:

- Step 1: identify strategic objectives
- Step 2: determine the roadmap's audience(s)
- Step 3: establish your roadmap's major themes
- Step 4: share created roadmap with relevant stakeholders
- Step 5: meet with your team to assign responsibilities

Case study setup: Case study works for over one yearfrom January to December 2018 withinan organization in the meal delivery industry. The basic information about the organization and case study setup are given in Table 1. We conducted quarterly observation sessions for agile team working in boxed two weeks sprints.

Total of 27 observation hours each cycle. The events observed were selected carefully to support our study with information on how the Scrum practices are implemented, observations on how the scrum team performing therequired activities like sprint review, retrospective and planning, understanding of how the major coordination events worked in practice, monitoring and measuring team velocity.

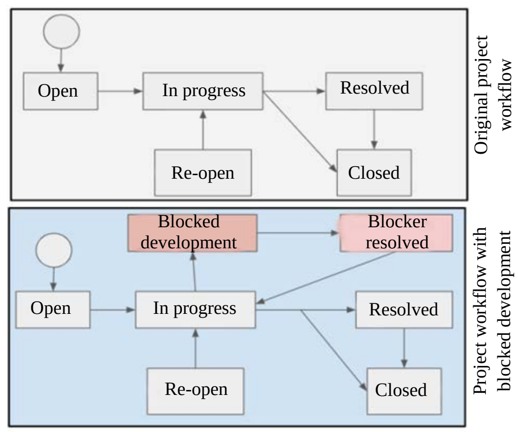

Fig. 2: Project workflow

Data collection: We have focused on two main sources for data collection along with ongoing observations, organization internal documentation ticketing system (JIRA) which is used for extracting and monitoring measurement metrics.

Baseline phase, a full cycle of three months focused on comprehensive organization data collection and analysis assummarized in Table 2, along with historical data for metrics we will use while monitoring our study. Study phase, three cycles of three months each, started by enriching the data warehouse team with the data steward role. Our case study applied for the data warehouse team, with 8-10 team members in 2018 with an average of 9 versus 3-7 for the same period the previous year (2017) with an average of 4.7 team members.

Starting the second cycle "blocked development" is introduced as a new ticket status to the workflow, in which the engineer can change ticket status to blocked development with adding reason in a comment, data steward was responsible for both planned tasks and any task marked as blocked development which he takes required actions to unblock dependency and return it to the responsible engineer. Data steward become part of scrum team, joining agile ceremonies. In Fig. 2, shows project workflow after introducing the new status.

By building a specialized data team in the organization, the focus was to get the best results out of this team, agile processes applied, however team productivity remains a question, during baseline phase, we have collected team productivity and translated it into velocity measurements to extract more insights about the data, then the data collection and analysis performed during three quarters following action plan methodology with one quarter evaluation period.

RESULTS AND DISCUSSION

We have focused on two main measures while observing data steward contribution and progress in the agile process. The first metric is the number of blocked work items, counting activities/work items prohibited completion due to dependence on other data related activity[41]. These blockers cannot be resolved by the individual assigned to complete the activity and the team needs data steward involvement and assistance to remove the blocker.

Average of 62% improvement in blocked work items ratio overall the three cycles. Using JIRA export, we have calculated blocked work items ratio to using:

- # resolved work items: aggregated count for unique "Issue key" where status in ("Done" and "Fixed") using month part of "Resolved" for aggregation
- # blocked work items: aggregated count for unique "Issue key" for all statuses except ("Won't do" and "Won't fix") using month part of "Blocked date" for aggregation

Blocked work items ratio = # blocked work items / # resolved work items. We have analyzed data over time for the three cycles after the baseline phase and built a trend analysis as displayed in Fig. 3.

The second measure used is productivity, represented by velocity in agile communities. Velocity is defined as the number of completed user stories in an iteration. It is often calculated as the sum of the number of story points required to complete a work item (206). We have used below calculations for measuring team velocity:

- Month: Month part of "Resolved" column. -# story points: as aggregated sum for "Custom field (Story Points)" column
- # team members: as unique count of "Assignee" column. Calculated fields
- Normalized story points = # story points/# team members

Velocity improvement ratio = (Normalized story points 2018-Normalized story points 2017)/Normalized story points 2017 (Table 3). In our study, we have compared team velocity for the case study period from April 2018 to December 2018 with team velocity for the same period previous year from April 2017 to December 2017. We have normalized numbers by the number of team members assigned to the tasks to make sure numbers are comparable. Also, normalizing resolve external factors impacting chart spikes, like Christmas time at end of the year reflects low velocity compared to other months as a lot of team members taking vacations. As detailed in Table 3, a total of 57.8% improvement in the number of story points delivered by each team member during our experiment.

Both measures showed a positive impact after introducing the data steward role to the team as it clarifies data quality, policies and procedures to team

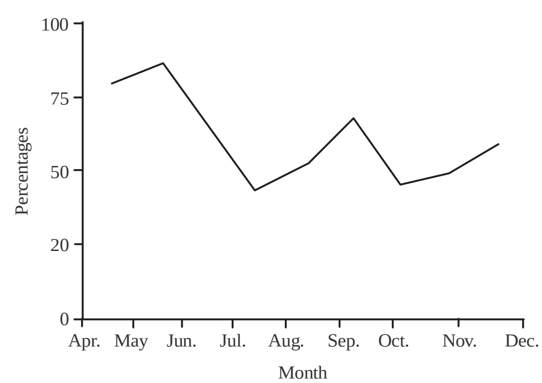

Fig. 3: Blocked work items ratio

Table 3: Story points results

Month	Story points 2017	Story points 2018	# Team members 2017	# Team members 2018	Normalized story points 2017	Normalized story points 2018
Apr.	75	105.5	6	10	12.5	10.6
May	55.4	85.5	6	10	9.2	8.6
Jun	33.8	33.5	7	9	4.8	3.7
Jul.	13.3	74.0	3	8	4.4	9.3
Aug.	3	179.5	3	10	1.0	18.0
Sep.	67.8	152.0	4	10	17.0	15.2
Oct.	21	103.5	4	10	5.3	10.4
Nov.	14.5	93.0	6	8	2.4	11.6
Dec.	17	82.5	4	6	4.3	13.8
Total	300.8	909.0	5	9	64	101.0
Improvement ratio					57.8%	

members during planning session and makes tasks more clear and ready for development, at same time steward handled many external meetings with different stakeholders ensure proper and clear communication.

CONCLUSION

The goal of our study was a contribution to maximize the chance of success for big data projects, hence, help organizations setting up their big data projects. We have focused on resolving data governance-related challenges through introducing data steward role to big data projects which run with agile scrum methodology. We have applied our proposed solution to one case study with a positive impact, so, we do recommend organizations start thinking about this important role within big data teams. Research case study was about applying research proposed solution through defining how the role of data steward can be involved to agile practices. Below key takeaways from our case study:

- With clear definition of his/her involvement during the process, we could keep the very strong role of data steward
- Involvement of data steward in daily standup discussion meetings resolved the issues earlier than at the end
- Introduction of “blocked development” status helps to determine blockers early during the sprint and resolve them with the responsible parties more effective
- Data steward data analysis results helped in understanding data we are dealing with and hence estimating more easily effort required for building product without need for adding extra effort for gaining skills and understanding data
- Data steward skills in identifying data quality make it easier for data engineers to focus on their related tasks
- The initial feedback workshop aligned team members including data steward about their goals, values and purposes and indicated the potential impact of resolving data governance topics on the success of data products and solutions within the organization

RECOMMENDATIONS

Further studies are needed to apply same idea in different business organizations and business domains and cultures to confirm the results obtained in our study. Such studies may lead to refinement of our proposed approach or creation of variations of it suitable for different environments and/or other agile methodologies.

REFERENCES

01. Reggio, G. and E. Astesiano, 2020. Big-data/analytics projects failure: A literature review. Proceedings of the 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), August 26-28, 2020, IEEE, Portoroz, Slovenia, pp: 246-255.
02. Kabanda, G., 2019. An evaluation of big data analytics projects and the project predictive analytics approach. Orient. J. Comput. Sci. Technol., 12: 132-146.
03. Goasduff, L., 2015. Gartner says business intelligence and analytics leaders must focus on mindsets and culture to kick start advanced analytics. Gartner IT Service Management Company, Stamford, Connecticut.
04. Hassani, R., Y.E.B. El Idrissi and A. Abouabdellah, 2017. Software project management in the era of digital transformation. Proceedings of the International Conference on Networked Systems, May 17-19, 2017, Springer, Marrakech, Morocco, pp: 391-395.
05. Tokuc, A.A., Z.E. Uran and A.T. Tekin, 2019. Management of Big Data Projects: PMI Approach for Success. In: Agile Approaches for Successfully Managing and Executing Projects in the Fourth Industrial Revolution, Bolat, H.B. and G.T. Temur (Eds.)., IGI Global, Pennsylvania, pp: 279-293.
06. Kumar, R., B. Subhash, M. Fatima and W. Mahmood, 2018. Quality assurance for data analytics. Qual. Assur., 9: 160-166.
07. Feki, M. and I. Boughzala, 2016. Big data governance: A literature review and research agenda. International Academic Association of Governance, France.
08. Al-Badi, A., A. Tarhini and A.I. Khan, 2018. Exploring big data governance frameworks. Procedia Comput. Sci., 141: 271-277.
09. Kim, H.Y. and J.S. Cho, 2018. Data governance framework for big data implementation with NPS case analysis in Korea. J. Bus. Retail Manage. Res., 12: 36-46.
10. Backscheider, N.E.A., 2015. Establishing data stewardship models. ECAR Working Group Paper, Louisville, Colorado.
11. Ballard, C., C. Compert, T. Jesionowski, I. Milman, B. Plants, B. Rosen and H. Smith, 2014. Information governance principles and practices for a big data landscape. International Business Machines Corporation, USA.
12. Teperek, M., M.J. Cruz, E. Verbakel, J.K. Bohmer and A. Dunning, 2018. Data stewardship-addressing disciplinary data management needs. Int. J. Digital Curation, 13: 141-149.

13. Grady, N.W., J.A. Payne and H. Parker, 2017. Agile big data analytics: AnalyticsOps for data science. Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), December 11-14, 2017, IEEE, Boston, Massachusetts, pp: 2331-2339.
14. Chen, H.M., R. Kazman and S. Haziyev, 2016. Agile big data analytics development: An architecture-centric approach. Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), January 5-8, 2016, IEEE, Kauai, Hawaii, pp: 5378-5387.
15. Dingsoyr, T., T. Dyba, M. Gjertsen, A.O. Jacobsen and T.E. Mathisen et al., 2019. Key lessons from tailoring agile methods for large-scale software development. IT Prof., 21: 34-41.
16. Frankova, P., M. Drahosova and P. Balco, 2016. Agile project management approach and its use in big data management. Procedia Comput. Sci., 83: 576-583.
17. Jurney, R., 2017. Agile data science 2.0, United States of America. O'Reilly Media, Inc., USA.
18. McAfee, A., E. Brynjolfsson, T.H. Davenport, D.J. Patil and D. Barton, 2012. Big data: The management revolution. Harvard Bus. Rev., 90: 60-68.
19. Zhou, Z.H., N.V. Chawla, Y. Jin and G.J. Williams, 2014. Big data opportunities and challenges: Discussions from data analytics perspectives (discussion forum). IEEE. Comput. Intell. Mag., 9: 62-74.
20. Shane, J., K. Strong, D. Gransberg and D. Jeong, 2015. Guide to project management strategies for complex projects (No. SHRP 2 Report S2-R10-RW-2). National Academies of Sciences, Engineering, and Medicine, Washington, USA.
21. Alharthi, A., V. Krotov and M. Bowman, 2017. Addressing barriers to big data. Bus. Horiz., 60: 285-292.
22. Douglas, M., 2013. Big data raises big questions. Government Technol., 26: 12-16.
23. Badrick, T., 2017. Big data, bigger opportunities. J. Lab. Precis. Med., 2: 43-43.
24. Weibl, J. and T. Hess, 2018. Success or failure of big data: Insights of managerial challenges from a technology assimilation perspective. Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI), March 6-9, 2018, Luneburg, Germany, pp: 47-58.
25. Schouten, P., 2013. Big data in health care: Solving provider revenue leakage with advanced analytics. Healthcare Financial Manage., 67: 40-43.
26. Small, M., 2019. Big data analytics-security and compliance challenges in 2019. KuppingerCole, Europe.
27. Sanger, J., C. Richthammer, S. Hassan and G. Pernul, 2014. Trust and big data: A roadmap for research. Proceedings of the 2014 25th International Workshop on Database and Expert Systems Applications, September 1-5, 2014, IEEE, Munich, Germany, pp: 278-282.
28. Zaidi, E., E. Thoo and N. Heudecker, 2019. Magic quadrant for data integration tools. Gartner Inc., Stamford, Connecticut.
29. Ionel, N., 2009. Agile software development methodologies: An overview of the current state of research. Annl. Univ. Oradea Econ. Sci. Ser., 18: 381-385.
30. Al-Marwaee, M.B.A.O., 2016. Techniques and strategies towards project management success. Int. J. Hybrid Inf. Technol., 9: 197-212.
31. Schmitz, K., 2018. A three cohort study of role-play instruction for agile project management. J. Inf. Syst. Educ., 29: 93-103.
32. Gharaibeh, N., I. Oti, D. Schrank and J. Zmud, 2017. Data management and governance practices. National Academies of Sciences, Engineering and Medicine, Washington, USA.
33. KISA., 2008. The importance of data governance and data quality management. Korea Internet and Security Agency, Korea.
34. Panian, Z., 2010. Some practical experiences in data governance. World Acad. Sci. Eng. Technol., 62: 939-946.
35. Plotkin, D., 2013. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance. Elsevier Science, Amsterdam, Netherlands,.
36. Wikipedia, 2020. Data steward. Wikimedia Foundation, Inc., San Francisco, California.
37. Peng, G., 2018. The state of assessing data stewardship maturity-an overview. Data Sci. J., Vol. 17, No. 7. 10.5334/dsj-2018-007
38. Shankarmani, R., R. Pawar, S.S. Mantha and V. Babu, 2012. Agile methodology adoption: Benefits and constraints. Int. J. Comput. Appl., 58: 31-33.
39. Anonymous, 2010. TDWI data governance fundamentals. The Data Warehousing Institute, Washington, USA.
40. Avison, D.E., F. Lau, M.D. Myers and P.A. Nielsen, 1999. Action research. Commun. ACM., 42: 94-97.
41. Anonymous, 2019. Agile metrics guide strategy considerations and sample metrics for agile development solutions. Department of Defense, USA.