11-29-2018

Identifying the Key Drivers for Teams to Use a Data Science Process Methodology

jeff saltz
syracuse university, jsaltz@syr.edu

IDENTIFYING THE KEY DRIVERS FOR TEAMS TO USE A DATA SCIENCE PROCESS METHODOLOGY

Research in Progress

Saltz, Jeffrey S, Syracuse University, United States, jsaltz@syr.edu

Abstract

While data science teams do not yet typically use a standard team process methodology, researchers are starting to explore process methodologies that improve team performance. However, little has been done to understand what might be the key acceptance factors for teams to implement a data science process methodology. To address this gap, the Diffusion of Innovation Theory is used as a theoretical lens to identify factors that might drive an organization to adopt a data science process methodology. The results of this qualitative research effort found ten factors that can influence a team to use, or not use, a data science process methodology. In short, eight positive factors were found with respect to relative advantage and compatibility and two negative factors were identified with respect to complexity. While more work is required to validate and refine these factors, the derived acceptance model can help teams as they consider adopting an improved data science process methodology.

Keywords: Data Science, Big Data, Project Management.

1 Introduction

Data science is an emerging discipline that combines expertise across a range of domains, including software development, data management and statistics. Data science projects typically have a goal to identify correlations and causal relationships, classify and predict events, identify patterns and anomalies, and infer probabilities, interest and sentiment (Das et al, 2015). With the increasing ability to collect, store and analyze an ever-growing diversity of data that is being generated with increasing frequency, the field of data science is growing rapidly. Because it is a new field, data science research typically has focused on improving data models and algorithms, but not on how to execute projects (Ahangama & Poo, 2015). In fact, in two literature reviews of previous research on the socio-technical issues when executing data science projects, little has been identified with respect to the processes and structures necessary to orchestrate those teams (Saltz and Shamshurin, 2016; Mikalef et al, 2017). However, studies have shown that the emphasis on technical issues such as tools, systems and skills have limited organizations from realizing the full potential of analytics (Ransbotham et al., 2015).

Furthermore, it has been observed that most data science projects are managed in an ad hoc fashion (Bhardwaj, 2015). Demonstrating the impact of this low level of process maturity for data projects, Kelly and Kaskade (2013) surveyed 300 companies, and reported that 55% of their data projects don't get completed, and many others fall short of their objectives. Hence, perhaps not surprisingly, it has been reported that an improved project management processes can result in higher quality outcomes for data projects (Mariscal et al, 2010; Saltz, Shamshurin & Crowston, 2017) and at least some managers are open to improving their process methodology, but might not think of doing it unless prompted (Saltz and Shamshurin, 2015).

Since the field is now starting to define project management methodologies for data science projects, there is a need to explore acceptance factors for using one of these data science process methodologies. Unfortunately, the factors driving the adoption of process methodologies in other domains might not be applicable to data science projects since Shin (2015) notes that traditional technology acceptance models require modifications in the context of new and emerging trends and technologies, such as the process used for data science projects. This is due to the fact that while data science projects have parallels to other domains, there are differences compared to these other types of projects. For example, compared to software development, data projects have a broader range of questions that could be addressed and an increased focus on data, such as what data is needed as well as the availability, quality and timeliness of the data (Dhar, 2013; Das et al, 2015; Saltz, 2015).

Finally, in exploring current challenges in creating value from business analytics Vidgen, Shaw and Grant (2017) note that one area of future research is in project management. Hence, to support this stream of future exploration, this article explores the acceptance factors for incorporating improved process methodologies and project management within a data science/analytics team. The rest of the paper is structured as follows. In Section 2, some additional background is presented, which is followed, in section 3 with a description of the methodology used in this study. Section 4 describes the findings from this research and section 5 presents a synthesis of the research results and also discusses limitations and possible next steps.

2 Background

While there has been some work on exploring the factors driving the acceptance of an organization to use big data analytics (Brock & Khan, 2017) and exploring the acceptance factors for the adoption of business intelligence & analytics within an organization (ex. Grublješič, Coelho & Jaklič, 2017), there has been minimal research on the acceptance factors for data science teams to use a data science process methodology. Perhaps the closest research that has been reported was an effort to understand the acceptance factors for using a capability and maturity model within a big data context (Saltz, 2017). Unfortunately, there have been no other research efforts identified on exploring the factors that could help understand the acceptance of a data science team process methodology. The rest of this section briefly explores what has been done to date with respect to data science process methodologies and then discusses a theoretical model to explore acceptance factors within a data science context.

2.1 Data Science Team Process Methodologies

While there has been some research on the challenges of doing data science, as previously mentioned, most of that research was focused on the technical challenges in executing the projects. Below, the recent efforts related to methodologies of teams working on data science projects are summarized. With respect to executing a data science project, current descriptions generally adopt a task-focused approach, conveying the techniques required to analyse data. For example, Jagadish et al (2014) describes a process that includes acquisition, information extraction and cleaning, data integration, modelling, analysis, interpretation and deployment. This step-by-step data science process description described does not provide much guidance about the process a data science team should use to work together (Saltz, 2015). Perhaps not surprisingly, it has been argued that projects need to focus on people, process and technology (Gao, Koronios & Selle, 2015; Grady, Underwood, Roy & Chang, 2014) and that task coordination is the main challenge for data projects (Espinosa & Armour, 2016). More broadly, Chen, Kazman & Kaziyev (2016) note how business analysts, data scientists, and other stakeholders need to work with system design, development, and operations teams to maximize value from big data is a key challenge. Furthermore, in exploring management challenges in creating value from analytics, Vidgen, Shaw and Grant (2017) found that team process issues are currently not top of mind, but they expect this situation to change as organizations advance in their usage of advanced analytics.

Hence, new methods for facilitating process integration and systematically achieving continuous value discovery and rapid deployment are needed (Chen, Kazman & Haziyev, 2016) and researchers have begun to address the need for a team-based data science process methodology via case studies to understand effective practices and success criteria (Das et al, 2015; Gao, Koronios & Selle, 2015; Saltz, Yilmazel & Yilmazel, 2016). In general, these studies have noted a lack of focus on process and team coordination, but when prompted, the data science teams were interested in improving their process.

Perhaps more interesting, in an experiment that explored different data science project management methodologies, Saltz, Shamshurin & Crowston (2017) reported on the effectiveness of several alternative methodologies. They identified that an Agile Kanban Data Science Methodology (AK-DSM), that combines a set of phases to do data science integrated with the pipeline process management from Kanban, was more effective than other agile methodologies or more waterfall like processes.

2.2 Theoretical Development of Acceptance Factors Model

From a theoretical perspective, the use of a data science project management process is a *process innovation*. Hence, the Diffusion of Innovation (DOI) Theory (Rogers, 1995) is an appropriate lens to examine the factors that could determine the assimilation of that process innovation. DOI defines innovation as "an idea, practice, or object that is perceived as new by an individual" and that DOI describes the factors that determine the assimilation (or adoption) of the innovation.

DOI has been extensively used to study information systems process innovation. For example, at one end of the spectrum, DOI has been used to examine IS process adoption over a period spanning four decades (Mustonen-Ollila & Lyytinen, 2003). More recently, DOI was leveraged to understand the acceptance factors for using a new agile software development technique (Schlauderer, Overhage & Fehrenbach, 2015). Since, as previously mentioned, data science projects differ from other disciplines such as software development in several key aspects, one cannot rely on the findings with respect to the specific acceptance factors identified in other contexts. However, one can note that the use of DOI for other information system processes supports the use in this context.

To summarize DOI, the theory states that the way potential adopters perceive the attributes of the innovation impacts the willingness of those individuals to assimilate (or adopt) that innovation. The five perceived attributes of the innovation are (Rogers, 1995; Moore & Benbasat, 1991):

- Relative advantage - the degree to which the innovation is perceived to be better than the idea it supersedes.
- Compatibility - the degree to which the innovation is perceived to be consistent with the existing values, past experiences and needs of potential adopters.
- Complexity - the degree to which the innovation is perceived to be difficult to understand and use.
- Observability - the degree to which the results of the innovation are visible to others.
- Trialability - the degree to which an innovation may be experimented with on a limited basis. This may include trying parts of a process or being able to watch others using the new process.

Previous research has noted some common impacts of these attributes (Schlauderer and Overhage, 2015). Specifically, the higher the perceived relative advantage, the more likely it is that the innovation will be adopted. In addition, if the process is amenable to being used on a limited bases, the trialability is also positively related to the process being adopted. However, if the innovation is perceived as an extreme change, then it will not be compatible with past experiences and is less likely to be adopted. In addition, if the process innovation is perceived as complex, it is also less likely to be adopted. Finally, if the observed effects are perceived to be small or non-existent, then this low observability reduces the likelihood of adoption.

3 Methodology

Case studies are particularly useful for in-depth studies of contemporary phenomena within the organizational context (Yin, 2003) in that the case study methodology provides better explanations and understandings of the examined phenomenon that would otherwise be lost in other quantitative designs (Miles & Huberman, 1994; Yin, 2003). Hence, an exploratory case study was conducted on teams using a process methodology to do a data science project. The case study focused on trying to understand the acceptance factors that might drive (or hinder) the adoption of a data science process methodology and hence, this research focused on the development of a data science project management acceptance model. In other words, the aim of the case study was to identify the acceptance determinants in the context of a data science project management process.

Since the use of data science process methodologies is not yet commonly used in industry (Saltz & Shamshurin, 2015), if data science teams in industry were surveyed, it was expected that most of the survey participants might not have a solid understanding of what it means to use a process methodology within a data science project. To address this potential concern, teams of graduate students were identified to participate in this study. One key advantage of using students was the ability to have each team use a data science team process methodology. Hence, the students were first required to use a data science process methodology. In this way, the students gained first hand knowledge of the key aspects of using a process methodology. The team members were then interviewed to identify potential acceptance factors. Therefore, when responding to the interview questions, students were able to leverage both their previous work experience, including the socio-technical situation within that organization, as well as leveraging the knowledge gained while they were graduate students participating in the case study.

Others have also used students to understand acceptance factors of information system related technology adoption. For example, Brock & Khan (2017) used graduate students to understand the acceptance factors of organizations thinking of using big data analytics. Further justifying the use of graduate students, as part of their analysis, Brock and Khan noted that there was “no reason to believe these students do not represent the average profile of IT professionals”. The rest of this section describes the participants in the study, the project and actual process used by the data science teams as well as the data collected during the case study.

3.1 Participants

In total, 92 people participated in the study. As previously noted, the participants were graduate students. All the students were part of a data science graduate level course, with the teams formed by the instructor. Eight-eight percent of the students were in an information management or applied data science graduate program and the remaining students were in programs ranging from business administration to public policy. Forty percent of the students were female and 80% had previous IT experience. Since most of the graduate students had previous experience working within an IT organization, they were well suited to use as research participants. The students were put into 18 teams, with 4 to 6 people per team. The teams all used the same process methodology, which is discussed later in this section. In addition, all the teams worked on the same data project.

3.2 Project Description

The group project started in the second week of the semester and continued until the end of the semester, thus lasting a total of twelve weeks. The project was done using the R programming language, a popular data science tool that is used in both industry and academia. The project teams were required to leverage many typical data science techniques, such as descriptive statistics, machine learning algorithms and geographic information analysis. The project was positioned in a way that the teams were to analyze a large data set of customer survey responses for a client. The dataset was a modified version of a real dataset of survey responses. Hence, the data was not real, but was representative of the actual challenges one might face in executing a data science project.

3.3 AK-DSM Process Description

All the teams used an Agile Kanban Data Science Methodology (AK-DSM). Saltz, Shamshurin & Crowston (2017) provide a high level description of AK-DSM. Below a brief description of the AK-DSM process used by the teams in this case study.

- *Visualize the workflow* – When using AK-DSM, the phases on the board included preparation (understand business context and the data), analyze (model/visualize, test/validate) and deploy (share/communicate results).
- *Limit Work In Progress* – Within each phase, there was defined a maximum number of work-in-progress tasks that could be in that phase (or column on the visual board).
- *Measure and manage flow* – As part of AK-DSM, teams were instructed to improve their process as they deemed appropriate. For example, by reviewing how work was flowing through the system, they could define new phases (columns on the board) or change WIP limits.
- *Make process policies explicit* – When using AK-DSM, the initial process was clearly defined, documented and explained to each team.
- *Improve collaboratively* – teams were encouraged to make small, incremental changes to the process to see the impact of those changes.

As one can see, the process leveraged key aspects of Kanban (Ahmad, Markkula & Oivo, 2013). Hence, when using AK-DSM, the team defined a prioritized list of what to do (via high level user stories, such as link weather data to our previously collected data). Then, based on the number of allowed simultaneous tasks at each phase, tasks flowed through the defined process. Limiting the number of tasks within any one step was a key focus to help to ensure the team minimizes bottlenecks, work in progress and to enable quick changes in team plans. Finally, as suggested by Saltz et al, students were given the freedom to refine the columns on their visual board, as their team thought best.

3.4 Data Collection and Analysis

Data collected during the case study included observations of the data science teams, a collection of project documents, and semi-structured interviews with project participants. However, most of the insight was gained via the analysis of the semi-structured interviews.

The semi-structured interviews focused on identifying the factors that determine the acceptance of using a data science process as well as perceptions of the benefit or detriment of those factors. Since a data science team process is a new concept for many data science teams, the questions were framed relative to executing projects with the "status quo" methodology, which was typically an ad hoc effort by the team to coordinate and communicate. The initial part of the interview focused on collecting general demographic information, such as their level of data science expertise as well as their overall IT experience. The rest of the survey focused on the factors that could determine the acceptance or rejection of their acceptance of a data science agile process methodology.

The information collected from the semi-structured interviews was used to identify specific factors that positively or negatively influence the acceptance of an agile process to guide data science teams. Specifically, the data collected in the interviews was analyzed in three steps (Miles and Huberman 1999). First, open coding was used to search the statements of the experts for recurring topics and to get an overview of the key points contained in the interviews. Second, the identified concepts were grouped by similar topics that were repeatedly articulated throughout the interviews. By using this process, it was possible to identify acceptance factors that were consistently mentioned in the interviews. In a third step, an applied theoretical coding was applied to categorize the identified acceptance factors according to the general acceptance factors defined when using DOI (Moore & Benbasat, 1991).

4 Findings

Based on the interviews, as shown in Figure 1, ten acceptance factors for using an agile data science team process methodology were identified. Below, the acceptance factors are described, grouped by the general acceptance factors defined in the DOI model.

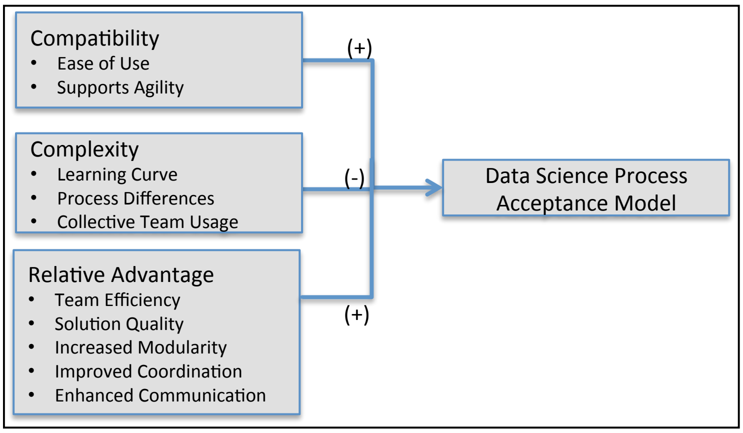

Figure 1: Data Science Process Methodology Adoption Model

4.1 Relative Advantage

There were five determinants identified relating to relative advantage (team efficiency, solution quality, increased modularity, improved coordination and enhanced communication). These are each described below.

Team Efficiency - One of the most common themes that emerged was the perception that, as one student noted, “there is a huge reduction of wasted work / wasted time”, or as noted by a different participant, the process was thought to “increase productivity and efficiency for the team”. Another student stated that process methodology enabled “a systematic approach to identify opportunities for improving efficiency” and “helped maximize our productive work”. One specific way that using the methodology was perceived to improved efficiency was that team members thought that it allowed for more effective workload allocation. For example, one observation was that it “helps in the ease of work distribution between the team members” and a different student noted that it “helps to divide the work and analyse the workflow that needs to be done”.

Solution Quality - The participants clearly perceived that their solution was better due to the use of AK-DSM and that was a key driver in the acceptance of using the process. This thought was summed up by one participant who noted that the process “allows a team to understand, implement and use unique solutions to develop various propositions”, or as stated by a different student, it was useful to “to invent and implement our own unique solutions”. Perhaps part of the driver for the solution quality was that the students thought that process helped them focus on the key aspects of the project. For example, “it helped us identify the key areas to focus on in our project” and “it helped us to overcome any issues/challenges faced in the project” were two comments supporting this line of thinking.

Increased Modularity - By forcing the teams to create tasks that get put on a visual board, team members were encouraged to think of working on the project in a more structured, more modular, manner. According to one team member, it “helps break down complex problem” and a different person stated that the "main advantage is that it helps you think about a unit work rather the whole solution" or more generally, a as noted by one student "it helps in delivering small portions of the bigger deliverable".

Improved Coordination - The use of the methodology also facilitated the team being organized and coordinated. For example, the process helped the team track work and distribute the workload among the team members. Demonstrating awareness of this, one person noted that "planning and execution was more visible, so we were able to stay on track with where we wanted and needed to be" and a different student stated that "this helped us identify the key areas to focus on in our project" and finally, another student commented that the process "helped us keep a track of the flow of our current progress while comparing our recent work with what we have already accomplished in the past, which led us to outline our goals for future" or simply that "it helped us track our tasks better and figure out the direction we needed to head in".

Enhanced Communication - Related to improved coordination, another determinate was improved communication within and across the team. This includes the team's ability to understand their project status and track their progress on their effort. Specifically, it was clear that the process made it much easier and much quicker for the team to understand the project's status. Many team members perceived that the use of the board made their project updates much easier and quicker to do. In this way, they perceived that the tool was useful for external communication. For example, one student noted that "project updates became easy, systematic and effective" and a different student commented "it's pretty clear to understand what we have done, what we are doing and what we are planning to do", and finally, a third student noted that "helped us articulate our progress in a more clear way".

4.2 Complexity

There were three determinants identified relating to complexity (learning curve, process differences and requires collective team usage). These were seen as negatively impacting the acceptance of the process methodology, and are each described below.

Learning Curve - One challenge that was noted was that the process, for some team members, took some time to learn. Hence, one negative determinate was the learning curve might be an issue. For example, one student noted that "the only disadvantage was the learning curve for the whole team to get on the same page" and a different student noted a similar thought in that "it took some time to get used to it. But after that it was quite easy".

Process Differences - Since the process was based on the Kanban methodology, the AK-DSM process was different than processes that many team members had previously used. This created situations where the team members wanted to revert back to using traditional project management tools. For example, by focusing on a visual board, but not on a Gantt chart, several people struggled that "there are no timeframes associated with each task, and it is difficult to predict the timeline until the items are completed". This suggests, perhaps, that part of the initial training should address some of these typical process differences.

Collective Team Usage - One road block to effective usage was that all the team members had to use the process. For example as one person stated "some team members did not use it regularly and hence the status was not updated frequently" and more broadly, a different student noted that "we didn't use it (the process) correctly, in that we used it to track what we had done as opposed to what we should be doing in the future." It is interesting to note that this last person understood what was expected – in terms of how to use the board, but due to team dynamics, the board was not used in the most effective manner. As a different example, some of the team members reported that they felt that the methodology was "not going to be useful, so why do the perceived extra work". This lack of use of the methodology frustrated the team members who wanted to fully use AK-DSM. This challenge exists with any methodology – if the entire team does not use it, the methodology is not useful for the subset of team members that try to use it. Additional training on the use and benefits of the methodology might encourage more students to more fully adopt the methodology.

4.3 Compatibility

There were two determinants identified relating to complexity (ease of use and supports agility). These are each described below.

Ease of Use - While some perceived that the process was difficult to learn, most stated that it was easy to use after they had gained an understanding of how to use it. For example, comments include statements such as "it is easy to use for beginners", "it was simple and outcome focused" and finally, it was "easy, intuitive and the systematic nature of the process makes it easy to track progress and monitor the recent updates".

Supports Agility - The process was perceived to support agility, including the ability to do continuous development. For example, one student noted that "it improves the delivery flow by promoting small, continuous changes in the system" and a different student noted that "the team was able to adjust the work in progress dynamically to avoid being idle" and finally, summing up the feedback of many students was high level the comment that it "gave us the speed and simplicity of agile development".

5 Discussion

A qualitative case study was used to identify and systemize the key determinants of an organization's acceptance of a data science process methodology. Based on the study's findings, a conceptual model of acceptance was proposed that leveraged the diffusion of innovation model of technology process acceptance. The case study focused on students who had previous industry experience.

While the case study explored the factors that influence the process adoption within a graduate course, the findings are relevant in the broader context of organizations doing data science. The research identified ten factors that could drive the adoption of a data science process methodology and the analysis suggests that there is a perceived relative advantage and compatibility in using an agile process methodology, but complexity (specifically a learning curve and a new process) were detriments to the use of the methodology. This acceptance model can help managers and data science leaders as they consider adopting an improved data science process methodology, which can lead to improved project outcomes.

5.1 Limitations and Future Research

One key limitation was that the case study was done with graduate students. While the students had previous industry experience, and the use of students enabled a large number of teams to be included in this research, the results might be different if data science teams in industry were surveyed. To address this limitation, a survey, leveraging this acceptance model, is a planned next step is to validate these findings.

In addition, the factors might vary by the type of organization, size of organization or by the type of big data project (Saltz, Shamshurin & Connors, 2017). The next phase of the research will also explore if some types of data science projects might have different acceptance models.

Finally, while the acceptance model was derived by teams using AK-DSM, many of the findings are likely applicable to data science teams contemplating using any enhanced team process. However, additional research is required to validate this applicability.

References

Ahangama S., Poo, D. (2015). "Improving Health Analytic Process through Project, Communication and Knowledge Management." *The International Conference on Information Systems (ICIS).*

Ahmad, M., Markkula, J., and Oivo, M. (2013). "Kanban in software development: A systematic literature review," in *Software Engineering and Advanced Applications* (SEAA), 2013 39th EUROMICRO Conference on.

Bhardwaj, A., Bhattacherjee, S., Chavan, A., Deshpande, A., Elmore, A., Madden, S., and Parameswaran, A. (2015). "DataHub: Collaborative Data Science & Dataset Version Management at Scale.*" Biennial Conference on Innovative Data Systems Research (CIDR).*

Brock, V., Khan, H. (2017). "Big data analytics: does organizational factor matters impact technology acceptance?" *Journal of Big Data*, 4(21).

Chen, H., Kazman, R., and Haziyev, S., (2016). "Agile Big Data Analytics for Web-based Systems: An Architecture centric Approach", *IEEE Transactions on Big Data*.

Das M., Cui R., Campbell D., Agrawal G., Ramnath R. (2015). "Towards Methods for Systematic Research On Big Data." *IEEE International Conference on Big Data*

Dhar, V. (2013). "Data science and prediction." *Communications of the ACM* 56.12.

Espinosa, J., Armour, F. (2016). "The Big Data Analytics Gold Rush: A Research Framework for Coordiation and Governance." *In: Proceedings of Hawaii International Conference on System Sciences (HICSS). IEEE.*

Gao J., Koronios A., Selle S. (2015). "Towards A Process View on Critical Success Factors in Big Data Analytics Projects." *Twenty-first Americas Conference on Information Systems (AMCIS)*

Grady N., Underwood M., Roy A., Chang W. (2014). "Big Data: Challenges, practices and technologies: NIST Big Data Public Working Group workshop." *IEEE International Conference on Big Data*

Grublješič, T., Coelho, P. S., & Jaklič, J. (2017). "The Shift to Socio-organizational Drivers of Business Intelligence & Analytics Acceptance". *Journal of Organizational and End User Computing*

Kelly, J., & Kaskade, J. (2013). "CIOS & BIG DATA What Your IT Team Wants You to Know." DOI= www.infochimps.com/resources/report-cios-big-data-what-your-it-team-wants-you-to-know-6/

Jagadish H., et al., (2014). "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86-94.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). "Big data analytics capabilities: a systematic literature review and research agenda", *Information Systems and e-Business Management*, 1-32.

Mariscal, G., Marban, O., and Fernandez, C. (2010). "A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies," *The Knowledge Engineering Review* (25:02)

Miles, M.B., and Huberman, A.M. (1999). *Qualitative Data Analysis*. London: Sage.

Moore, G., and Benbasat, I. (1991). "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation." *Information Systems Research* (2:3)

Mustonen-Ollila, E., and Lyytinen, K. (2003). "Why Organizations Adopt Information Systems Process Innovations: A Longitudinal Study using Diffusion of Innovation Theory." *Information Systems Journal* (13:3), pp. 275-297.

Ransbotham, S., Kiron, D., and Prentice, P.K. (2015). "Minding the Analytics Gap." *MIT Sloan Management Review* (56:3), pp. 62-68.

Rogers, E.M. (1995). *Diffusion of Innovations* , (4. ed.). New York, NY: Free Press.

Saltz, J. (2017). "Acceptance Factors for Using a Big Data Capability and maturity Model", *In Proceedings of the 25th European Conference on Information Systems (ECIS).*

Saltz, J., Yılmazel, O., Yılmazel, S. (2016). "Not all software engineers Can Become Good Data Engineers", in *Big Data (Big Data), 2016 IEEE International Conference on*.

Saltz J., Shamshurin I., and Crowston K. (2017). "Comparing Data Science Project Management Methodologies via a Controlled Experiment." *Hawaii International Conference on System Sciences (HICSS)*

Saltz, J., Shamshurin, I., Connors, C., (2017). "Predicting Data Science Socio-Technical Execution Challenges by Categorizing Data Science Projects", *Journal of the Association for Information Science and Technology.*

Saltz, J., Shamshurin, I. (2016). "Big Data Team Process Methodology: A Literature Review and the Identification of Critical Factors for a Project's Success." In *Big Data (Big Data), 2016 IEEE International Conference on*.

Saltz, J., Shamshurin, I. (2015). "Exploring the process of doing data science via an ethnographic study of a media advertising company." *IEEE International Conference on Big Data*

Saltz, J. (2015). "The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness." *IEEE International Conference on Big Data*

Schlauderer, S., Overhage, S., & Fehrenbach, B. (2015). "Widely Used but also Highly Valued? Acceptance Factors and Their Perceptions in Water-Scrum-Fall Projects." *The International Conference on Information Systems (ICIS).*

Shin, D.H. (2015). Demystifying big data: Anatomy of big data developmental process. Telecommunications Policy, 40(9), 837–854

Vidgen, R., Shaw, S., & Grant, D. B. (2017). "Management challenges in creating value from business analytics". *European Journal of Operational Research*, 261(2), 626-639.

Yin, R.K. (2003). "Case Study Research." *Design and Methods*. Thousand Oaks, CA: Sage Publications.