Spring 6-10-2017

ACCEPTANCE FACTORS FOR USING A BIG DATA CAPABILITY AND MATURITY MODEL

Jeffrey Saltz
Syracuse University, jsaltz@syr.edu

ACCEPTANCE FACTORS FOR USING A BIG DATA CAPABILITY AND MATURITY MODEL

Research in Progress

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

Abstract

Big data is an emerging field that combines expertise across a range of domains, including software development, data management and statistics. However, it has been shown that big data projects suffer because they often operate at a low level of process maturity. To help address this gap, the Diffusion of Innovation Theory is used as a theoretical lens to identify factors that might drive an organization to try and improve their process maturity. Specifically, thirteen acceptance factors for teams to use (or not use) a Big Data CMM are identified. These results suggest that a positive perception exists with respect to relative advantage, compatibility and observability factors, and a negative perception exists with respect to perceived complexity. While more work is required to refine the list of factors, this insight can help guide the improvement of big data team processes.

Keywords: Data Science, Big Data, Project Management.

1 Introduction

A big data project is one that uses statistical and machine-learning techniques on large volumes of unstructured and/or structured data generated by systems, people, sensors or digital traces of information from people. This work is done in a distributed computing environment with 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). These projects promise automated actionable knowledge creation and predictive models for use by both humans and computers and makes it feasible for a machine to ask and validate interesting questions humans might not consider (Dhar, 2013).

Because it is a new field, big data research typically has focused on improving data models and algorithms, but not on how to execute projects (Ahangama & Poo, 2015; Saltz & Shamsurin, 2016). However, studies have shown that this biased emphasis on technical issues such as tools, systems and skills have limited organizations from realizing the full potential of analytics (Ransbotham et al., 2015). Indeed, it has been argued that projects need to focus on people, process and technology (Gao et al, 2015; Grady et al, 2014) and that task coordination is the main challenge for data projects (Espinosa and Armour, 2016). Furthermore, it has been observed that most big data projects are managed in an ad hoc fashion, that is, at a low level of process maturity (Bhardwaj, 2015). Demonstrating the impact of this low level of process maturity for big data projects, Kelly and Kaskade (2013) surveyed 300 companies, and reported that 55% of big data projects don't get completed, and many others fall short of their objectives. While there are many reasons a project might not get completed, with a robust team-based process methodology, one would expect many of those reasons to be identified prior to the start of the project, or to be mitigated via some aspect of the project execution and/or coordination methodology. Perhaps not surprisingly, it has also been reported that an improved process model would result in higher quality outcomes (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).

Maturity models are tools that facilitate the assessment of the level of development of organizational capabilities, processes, and resources (Cosic et al., 2012). One key goal of using a maturity model is to drive process improvement. In fact, the Capability Maturity Model (CMM) has been one of the most popular process improvement approaches used to improve the quality of software products (Unterkalmsteiner, 2012) and many studies have demonstrated that that investments in CMM process improvement have delivered enhanced software quality, reduced software development costs as well as other more intangible benefits (Krishnan et al, 2000; Harter et al, 2000; Diaz & Sligo, 1997; Ferreira et al, 2008; Hyde & Wilson, 2004). One reason for this improvement is that it is generally accepted that software processes need to be continuously assessed and improved in order to fulfill the requirements of the customers and stakeholders of the organization (Fuggetta, 2000).

With this in mind, one possible way to improve the process teams use to execute big data projects is to adopt a Capability and Maturity Model for big data projects (BD-CMM). However, CMM has not yet been applied within a big data context. The factors driving the adoption of maturity models in other domains might not be relevant to big data projects. This is due to the fact that while big data projects have parallels to other domains, there are differences as compared to these other types of projects. For example, compared to software development, big data projects have a broader range of questions that could be addressed and an increased focus on data, including trying to determine what data is needed and the availability, quality and timeliness of the data (Saltz, 2015; Dhar, 2013; Das et al, 2015). In addition, since big data projects are typically evolutionary and experimental nature, using a CMM could constrain (or perceive to constrain) big data projects and adversely affect their success. This suggests that the factors driving the adoption of a more mature project methodology within a big data context might be different from the factors identified in other domains. Furthermore, after performing an extensive literature review, Poeppelbuss, Neihaves, Simons & Becker (2011) noted that theories on the adoption of maturity models within Information Systems is distinctly rare.

Hence, this research effort aims to explore the acceptance or rejection factors for a BD-CMM, specifically focusing on the following research questions:

What are the factors that will determine the acceptance or rejection of a BD-CMM?

Compared to a team's current process, are these factors perceived as a benefit or detriment?

In selecting a theoretical foundation for this research, it can be noted that introducing and using a capability and maturity model within an organization would be a process innovation to that organization. As such, the diffusion of innovation (DOI) theory (Rogers, 1995) is an appropriate lens upon which to frame this research, in that 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.

In the following section, a review of the literature related to the use of CMM within a big data context as well as exploring research on the factors that have driven the adoption of CMM within other domains is provided. Next is a description of the theoretical framework used to conduct the study, which is followed by the methods section. The findings are then discussed and the paper concludes with a discussion of the expected contribution as well as the implications and limitations of this research.

2 Literature Review

While there has been no research reported on factors that could determine the acceptance of a BD-CMM, below previous research on using a capability and maturity model within a big data context and CMM acceptance factors in other domains are summarized.

2.1 Big Data CMM

Many models for big data organizational readiness have been created to help organizations understand impediments to leveraging big data. For example, IBM's maturity model (Malik, 2013) provides broad-based organizational readiness dimensions. This is similar to IDC's maturity model (Vesset, 2013), which speaks to the organization's ability to successfully execute big data projects. These organizational maturity models are helpful to understand if there are fundamental impediments that are independent of the big data team that executes the data project. For example, IBM defined five components (Business Strategy, Information, Analytics, Culture and Execution, Architecture, Governance) for each level of their five levels of maturity (Ad hoc, Foundational, Competitive, Differentiating, Breakaway). There are minimum requirements for each of these components of maturity for each level of organizational maturity. IDC has defined the most mature framework (known as the "big data and Analytics MaturityScape"). This framework consists of five stages of maturity, which essentially parallel IBM's (Ad hoc, Opportunistic, Repeatable, Managed, and Optimized). Similar to IBM's model, IDC's model is measured across five components (intent, data, technology, process, and people). Note that the 'process' noted by IDC is not the maturity of the process used by the big data team, but rather, notes the process the organization uses to leverage data insight (i.e., not generating data insight) and includes items such as access to siloed information, data analytics vs data tracking, and documenting the decision process.

In addition to the industry defined organizational readiness models, there has also been a bit of research on maturity models that focus on big data organizational readiness. One example looked at a maturity model for a specific big data domain (Sulaiman et al, 2015). While useful, there was no focus on what might be the factors that would drive people to adopt the model, and equally important, the focus was not on the process the team should use to perform a big data project, but rather, on the organizational readiness of the entire organization to use big data analytics. In a different example, Moore (2014) discussed big data maturity models, and mentioned IDC's model as one of the most applicable. While Moore also mentioned the broadly used Capability and Maturity Model (Paulk, Curtis, Chrissis & Webber, 1993), there was no discussion on how CMM could be used for big data, nor how CMM related to IDC's organizational readiness assessment. Yet another effort developed a Community Capability Model Framework (CCMF) designed to assist institutions and researchers to enhance the capability of their communities to perform data-intensive research (Lyon et al, 2012). This research explored the rationale for using capability modeling and outlined the main capabilities underlying the current version of the CCMF, but did not discuss what factors might drive the adoption of using CCMF. Finally, Crowston and Qin (2011) suggested a model that defined the key process areas and practices necessary for effective scientific data management (SDM). While this is useful to help understand the process maturity for teams doing SDM, the article did not discuss the factors that might encourage or discourage the use of the maturity model.

2.2 CMM Acceptance Factors in other domains

As previously noted, big data projects are different from software engineering projects. However, it still might be helpful to understand what has been researched with respect to the acceptance factors for adopting a CMM-focused Software Process Improvement (SPI). Unfortunately, research on theories of the adoption of maturity models within Information Systems is rare. However, three studies are very relevant. In one report, Staples and Niazi (2008) investigated why organizations adopt CMM-based SPI approaches. After examining over forty studies, they found that, independent of organizational size, CMM-based SPI has mostly been adopted to help organizations improve project performance and product quality issues. Taking a different approach, two other reports explored why organizations do not adopt CMM (Staples et al, 2007; Khurshid, Bannerman & Staples, 2009). The most frequent reasons given for not adopting a CMM-focused process improvement effort was that it was perceived to be too costly, there was uncertainty with respect to benefits, or that the organization had no time to implement the initiative.

3 Theoretical Development

It has been noted that using a maturity model could help that organization improve their processes. Hence, the use of a CMM for big data efforts could be thought of as a process improvement / innovation for that organization. Therefore, from a theoretical perspective, one can classify the use of a CMM to develop an improved big data project management process as a *process innovation*. Hence, the Diffusion of Innovation (DOI) Theory would be an appropriate lens to examine the factors that could determine the assimilation of that process innovation.

DOI has been extensively used to study information systems process innovation. For example, at one end of the spectrum, DOI was 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, big data 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. The higher the perceived advantage, the more likely the innovation will be adopted.
- Compatibility— the degree to which the innovation is perceived to be consistent with the existing values, past experiences and needs of potential adopters. 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.
- Complexity— the degree to which the innovation is perceived to be difficult to understand and use. Innovations that are perceived as complex are less likely to be adopted.
- Observability— the degree to which the results of the innovation are visible to others. If the observed effects are perceived to be small or non-existent, then the likelihood of adoption is reduced.
- Trialability—the degree to which an innovation may be experimented with on a limited basis. This may include trying out parts of a program or having the opportunity to watch others using a new program. Trialability is positively related to the likelihood of adoption.

4 Methodology

Since the focus of this effort was on the factors that can drive the start of an effort to improve a big data team's process maturity, as opposed to the challenges of improving a group's big data process maturity or understanding a big data group's maturity, this research did not observe a big data team trying to enhance their maturity, but rather, leveraged semi-structured interviews with big data practitioners to understand the acceptance factors that might drive (or hinder) the adoption of an effort to improve an organization's BD-CMM. In other words, the research was focused on what might drive an organization to improve their big data process maturity.

Hence, the interviews focused on identifying the factors that determine the acceptance of using a BD-CMM as well as perceptions of the benefit or detriment of those factors. Since a BD-CMM is a new concept for many big data teams, these questions were framed relative to executing projects with the "status quo" of continuing to perform projects with their current process methodology as compared to working to improve the way their team works together to execute projects.

Due to the fact that there is currently not a lot of knowledge about a big data team's perceptions of improving their process methodology and using a BD-CMM, part of the interview was a more open-ended discussion with the practitioners to identify possible factors that could impact an organizations ability to successfully execute big data projects. The use of a semi-structured interview was leveraged since semi-structured interviews provides both breadth and depth of the discussions, while also enabling the comparison of results between the discussions by following a common outline of questions (Yin, 2003). These interviews, which followed the structure described by Myers and Newman (2007), focused on gaining an understanding of the factors that might influence the acceptance or rejection of using a big data maturity model. To better ensure the reliability of the results and to examine perceptions from multiple perspectives, people were interviewed with different roles across multiple organizations. Since these roles (such as project managers and data scientists) were experts in their field, and that domain experts typically offer significant insights of the desired domain, the number of interviewees can be low (Bogner, Littig & Menz, 2009).

The initial part of the interview focused on collecting general demographic information, such as their level of big data expertise as well as the experience the interviewee had managing big data projects. Then the interviewer provided some context for process improvement and the use of a capability and maturity model. The rest of the interview focused on the factors that could determine the acceptance or rejection of a process improvement effort, and specifically, using a capability and maturity model for improving how a team executes big data projects. The research questions were based on the DOI Theory and covered the perceived advantages and disadvantages in comparison to their current practices. In addition, other questions focused on which factors were perceived to be more or less compatible with the way their current team and organization worked and which factors increased or reduced the complexity of a big data project. Finally, factors relating to the observability and trialability of using a maturity model were explored. Each interview focused on identifying possible factors that might be applicable within that person's organization as well as other factors that might be applicable in general, even if not useful for that specific organization. The questions were structured in this way to facilitate brainstorming what factors might be useful, and not forcing the interviewees to focus on how that factor might or might not be appropriate within their current project. Finally, the discussions also covered how each factor influenced the expert's desire to use the process innovation within their organization.

The information collected from the semi-structured interviews was used to identify specific factors that positively or negatively influence the acceptance of a big data maturity model. The data collected in the interviews was analyzed in three steps (Miles and Huberman 1999). First, an 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.

5 Findings

To better ensure the reliability of the research results and to examine perceptions from multiple perspectives, people were included with different roles across multiple organizations. In total, ten people were interviewed from three different organizations.

One organization was a large Fortune 500 company in the financial services industry, one was a large organization in the insurance industry ramping up their big data efforts, and one was a smaller organization in the entertainment sector. While all the organizations had active big data efforts underway, not surprisingly, the two largest organizations had multiple teams executing multiple projects while the smallest organization had only one team working on one main project.

Of the ten people that were interviewed, five were from the large financial services organization, two were from the organization in the insurance sector and three were from the smallest organization. The roles of interviewees included two clients (internal to the organization), four big data managers, three data scientists and one project manager (in the financial services organization). The reason only one project manager was interviewed was that the other two organizations did not have explicit project manager roles within the big data efforts, but rather, within those two organizations, the big data managers typically acted as the project manager.

Half of the interviewees had significant (10+ years) experience with data analytic focused projects, and thirty percent had significant project management experience. None of the interviewees were using a capability and maturity model, but sixty percent had knowledge of capability and maturity models, within the context of software development.

Based on the interviews, as shown in Figure 1, 13 acceptance factors for using a BD-CMM were identified. Since this list was generated via semi-structured discussions, and it was possible some people didn't think of all the key factors, the threshold was kept low and documented acceptance factors that were noted by more than one third of the interviewees. This low threshold also takes into account the potential differences in roles within the big data team, where an acceptance factor might be important for one role, but not others. Below, the acceptance factors are described, grouped by the general acceptance factors defined in the DOI model.

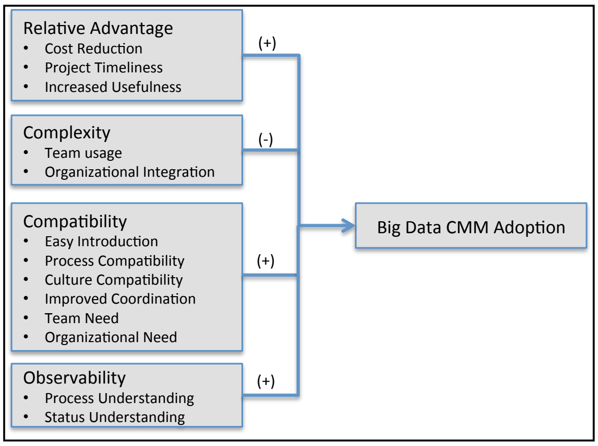

Figure 1: BD-CMM Adoption Model

5.1 Relative Advantage

There were three factors relating to relative advantage. The first of these was ***cost reduction***, which suggests that using a BD-CMM might drive down the costs of executing big data projects. For example, a client exemplified this view by stating "we currently waste a lot of time arguing about data requirements at the start of the project. If we had a better process, perhaps more iterative, it could be much faster for me to work with the data and IT teams. In order to do this, we need a process improvement methodology and I do think this would reduce costs and time in the long run". The second relative advantage factor was ***project timeliness***. This factor captures the belief that a BD-CMM might enable teams to deliver projects in a timelier manner. For example, a manager noted that "we often have to wait for technology infrastructure. If we could improve our process to the point where that wasn't an issue, and I don't think it would be an issue, then that would be reason enough to work on process improvement". Finally, ***increased usefulness*** refers to the perception that by better understanding the client's needs, the results could be more useful to the project champion (ex. client) perhaps by better understanding what the client needs or perhaps by the team generating more useful analytics. This line of thinking is summarized by one manager's quote "If we are all better coordinated, then I think we could do more insightful analytics. For example, sometimes we figure out we need data too late in the process". In another example, a different manager noted that "If we can make sure we are all on the same page, with respect to what are the requirements, or more appropriately, desires, of our client, that would be great" and that "Sometimes, it is difficult to really understand what the business wants – not just a wish list of everything possible, but a realistic, prioritized list. I think an improved process would help everyone understand the importance of this fact".

5.2 Complexity

There were two factors relating to complexity. The first, ***team usage,*** focuses on the fact that using a BD-CMM might be perceived to be complex to use within the team. For example, one data scientist noted that he was "not sure how this would actually work within our team, since it seems like it's just adding work". The second factor, ***organizational integration,*** describes the situation where using a BD-CMM is perceived to be difficult to integrate within organization. One manager, who had significant software management experience, noted that "I previously worked for an organization where they wanted to improve how we worked, but we could never discuss CMM or getting to a certain maturity level. In other words, the culture was very much against getting to a specific maturity level just for the sake of getting to a specific maturity level". Hence, it is possible that some people will think using a BD-CMM will just be process improvement for the sake of process improvement (i.e., no benefits).

5.3 Compatibility

There were six factors identified relating to compatibility. The first, ***easy introduction***, notes that for some organizations, it will be easy to adopt a BD-CMM. As one client stated, "since we are new to data analytics, I'd support any process you thought made sense", and a data scientist mentioned a similar thought "sure why not, everyone is open to doing things better. We just haven't given the process we use much thought". ***Process compatibility*** addresses the fact that for some organizations, using a BD-CMM would be consistent (or inconsistent) with existing processes used within the organization. For example, one data scientist noted that "we already have a great deal of focus on process – so I think people would be very open to process improvement". The third compatibility factor is ***culture compatibility***, suggesting that for some organizations, using a BD-CMM will be consistent with existing organizational culture. "We already have to integrate our data analytics effort into an IT SDLC methodology, so I think they would understand if we developed a specific data analytics process, and process improvement effort" noted the project manager. A fourth factor, ***improved coordination***, notes that using a BD-CMM could address an organizational challenge of uncoordinated efforts within the project, typically with extended team members. "Anything that improves coordination between all the groups working on this project would be a huge win, I would support it", noted the project manager. The fifth factor, ***team need***, is related to the fourth in that both address team needs. However the fifth factor is more general. For example, establishing reliable project timelines are often a challenge, as noted by a senior data scientist "Sometimes, as a team, we do not create realistic project timelines. Not sure how to solve this, but there must be a way to define an agile-like process where it should be easier to estimate task completion". Finally, the sixth factor is **organizational need**, where using a BD-CMM could address an issue within the organization, such as one noted by a different data scientist "We could provide good analytical solutions, but many in the organization do not want to take the time to understand. If we had an improved process, where the extended team was more engaged, then I think everyone would better understand the possibilities of what could be achieved".

5.4 Observability

There were two factors identified relating to observability. The first, ***process understanding***, notes that using a BD-CMM can enable extended team members to have an improved understanding of the process used by the team. For example, a manager stated that "It takes me a lot of time to get everyone on the same page, anything that makes that task easier is great". The second factor, ***status understanding***, captures the perceived belief that a BD-CMM can address the need that clients and managers sometimes do not have an accurate and timely understanding of the project's status. For example, a client noted that “If I could get a better understanding of current status and issues that would be great – today, it just seems like everything is a surprise, which leads me to think that things are out of control”.

5.5 Trialability

The model has no factors identified with trialability. It is perhaps surprising that trialability was not mentioned, and when prompted, was not thought of an important factor. Given the experimental nature of big data, this might be considered a counter intuitive finding. One might expect people to want to try a big data proof-of-concept project. However, perhaps due to the fact that the organizations surveyed already had invested in big data, the challenges were perceived to be more focused on how to actually execute a big data project, as compared to trying to prove the viability of a big data in general.

6 Conclusion

The focus of this research is to understand the factors that can drive the adoption of using a Big Data CMM. One interesting finding is that the identified factors were similar across the roles within the organizations surveyed (clients, managers, data scientists and project managers). In addition, the initial results, across two key industries (financial services and entertainment), suggest that the findings are not industry specific. However, more people, across more diverse set of industries need to be surveyed to validate these findings. To address this limitation, in terms of the diversity and number organizations, one planned next step is to validate these findings via a more broad-based survey. This future research will explore if these factors vary by the type of organization, size of organization or by the type of big data project (Saltz, Shamshurin & Connors, 2016). This next phase of the research will also explore if some of the factors are significantly more important than other factors.

This research identified thirteen factors that could drive the adoption of a big data CMM. The results suggest that there is a perceived relative advantage of using a big data CMM in terms of project cost, project duration and having the results be more useful to the project champions. In many situations, it was also perceived to be compatible with the current organization in terms of addressing a perceived need, improving coordination, and being compatible in terms of the culture of the team and organization. Observability was identified as a factor in terms of enabling the broader team to understand the process being used as well as enabling that extended team to understand the status of the effort. However, not surprisingly, complexity was viewed as a negative factor, both in terms of team usage and integration within the organization.

References

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

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).*

Bogner, A., Littig, B., and Menz, W. (2009). *Interviewing Experts*. Palgrave Macmillan.

Cosic, R., Shanks, G., and Maynard, S. (2012). "Towards a Business Analytics Capability Maturity Model." in *Proceedings of the 23rd Australian Conference on Information Systems*, J. Lamp (ed.), Deakin University, Geelong, Victoria.

Crowston, K., & Qin, J. (2011). "A capability maturity model for scientific data management: Evidence from the literature." *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1-9.

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.

Diaz M, Sligo J. (1997). "How software process improvement helped Motorola." *IEEE Software*; 14(5):75–81.

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.*

Ferreira, A., Santos, G., Cerqueira, R., Montoni, M., Barreto, A., Rocha, A., Barreto, A., and Silva, R. (2008). "ROI of software process improvement at BL Informatica: SPIdex is really worth it." *Software Process Improvement and Practice*, vol. 13, no. 4, pp. 311–318.

Fuggetta, A. (2000). "Software process: a roadmap". In *Proceedings Conference on The Future of Software Engineering*, Limerick, Ireland, pp. 25–34.

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*

Harter, D., Krishnan, M., and Slaughter, S. (2000). "Effects of Process Maturity on Quality, Cycle Time and Effort in Software Product Development." *Management Science*, vol. 46, pp. 451-466.

Hyde K., & Wilson, D. (2004). "Intangible benefits of CMM-based software process improvement." *Software Process Improvement and Practice*, vol. 9, no. 4, pp. 217–28.

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/

Khurshid, N., Bannerman, P. L., & Staples, M. (2009). "Overcoming the first hurdle: Why organizations do not adopt CMMI." In *Trustworthy Software Development Processes* (pp. 38-49). Springer Berlin Heidelberg.

Krishnan, M., Kriebel, C., Kekre, S., Mukhopadhyay, T. (2000). "An Empirical Analysis of Productivity and Quality in Software Products." *Management Science*, vol. 46, no. 6.

Lyon, L., Ball, A., Duke, M., & Day, M. (2012). "Developing a Community Capability Model Framework for data-intensive research." In *Proceedings of the 9th International Conference on the Preservation of Digital Objects* (pp. 9-16).

Malik, P. (2013). "Governing big data: principles and practices." *IBM Journal of Research and Development*, *57*(3/4), 1-1.

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, D. (2014). "Roadmaps and Maturity Models: Pathways toward Adopting Big Data." *Proceedings of the conference for Information Systems Applied Research.*

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.

Myers, M.D. and Newman, M. (2007). "The Qualitative Interview in IS Research: Examining the Craft." *Information and Organization* (17:1), pp. 2-26.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. (1993). "Capability maturity model Version 1.1" *IEEE Software*, 10(4), pp 18-27.

Poeppelbuss, J., Niehaves, B., Simons, A., & Becker, J. (2011). "Maturity models in information systems research: literature search and analysis." *Communications of the Association for Information Systems*, *29*(27), 505-532.

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., 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., (2016). "A Framework for Describing Big Data Projects." In: *19th International Conference on Business Information Systems.*

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).*

Staples, M. & Niazi, M. (2008). "Systematic review of organizational motivations for adopting CMM-based SPI." *Information and software technology*, *50*(7), 605-620.

Staples, M., Niazi, M. Jeffery, R., Abrahams, A., Byatt, P., & Murphy, R. (2007). "An exploratory study of why organizations do not adopt CMMI." *Journal of systems and software*, *80*(6), 883-895.

Sulaiman, H., Cob, Z. C., & Ali, N. A. (2015). "Big data maturity model for Malaysian zakat institutions to embark on big data initiatives." *Software Engineering and Computer Systems (ICSECS*).

Unterkalmsteiner M, Gorschek T, Islam AM, Cheng CK, Permadi RB, Feldt R. (2012). "Evaluation and measurement of software process improvement—a systematic literature review." *Software Engineering*, IEEE Transactions on; 38(2):398–424.

Vesset, D., Versace, M., Gerard, G., O'Brien, A., Burghard, C., Feblowitz, J., Osswald, D., Ellis, S. (2013). "IDC MaturityScape: Big Data and Analytics – A Guide to Unlocking Information Assets." IDC.

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