Big Data Team Process Methodologies: A Literature Review and the Identification of Key Factors for a Project's Success

Jeffrey S. Saltz
Syracuse University
Syracuse, NY USA
jsaltz@syr.edu

Ivan Shamshurin
Syracuse University
Syracuse, NY USA
ishamshurin@syr.edu

Abstract — **This paper reports on our review of published research relating to how teams work together to execute Big Data projects. Our findings suggest that there is no agreed upon standard for executing these projects but that there is a growing research focus in this area and that an improved process methodology would be useful. In addition, our synthesis also provides useful suggestions to help practitioners execute their projects, specifically our identified list of 33 important success factors for executing Big Data efforts, which are grouped by our six identified characteristics of a mature Big Data organization.**

Keywords- Analytics Process, Big Data, Data Science, Project Management, Process Methodology

I. INTRODUCTION

Current Big Data research has typically focused on improving data models and algorithms, but not on understanding the best approach to execute projects [1] [2]. Furthermore, it has been noted most teams that do big data work in an ad hoc fashion, using trial and error to identify the right tools, that is, at a low level of process maturity [3]. Hence, not surprisingly, it has been reported that an improved process model would result in higher quality outcomes [4]. Furthermore, it has also been reported that a project management methodology from a related field, such as software development or operations research, fails to address specific big data challenges [2]. Hence, there is a strong need have a methodology specifically for Big Data projects. Among the reasons for needing a Big Data specific methodology are that big data projects are often exploratory in nature, have business requirements that are often are not clearly specified, and that the results can be challenging to validate [2].

However, there has been little reported on potential team process methodologies that could be used for Big Data projects, and what has been reported is scattered across numerous conferences and journals. Due to this, there is no coherent review and analysis of the work that has explored big data project methodologies. To address this gap, this paper reports on a review of previously publications with respect to how teams execute big data projects.

Specifically, we assessed 10,028 Big Data (or data science) related papers published at recent conferences and journals. The aim of our study was to synthesize the currently reported research on process methodologies for teams executing Big Data projects. Specifically, we focused on two research questions:

RQ1. What are the current approaches to plan, organize and perform big data projects?

RQ2. What are the issues and actionable insights that we can synthesize from our literature analysis?

II. BACKGROUND

Existing approaches for describing how to execute a data project have often focused on the steps a team should take, such as preparation, analysis, reflection, and dissemination. These models include CRISP-DM [5] and SEMMA [6]. However, their usage is decreasing [7] [8], in part because these approaches do not provide a view on communication management, knowledge management and project management [1]. Related to this, Espinosa and Armour [9] noted that the main challenge for data projects is task coordination and a recent Gartner report advocates for more careful management of analysis processes, though a specific methodology is not identified [10].

Without a robust methodology, teams are overly dependent on the experience of their senior data analytic leaders. This lack of robust team-based methodology is similar to how software was developed in the late 1960's, which caused significant issues during the first wave of software development [11]. For example, it has been reported that three quarters of corporate business intelligence projects fail due to poor communication [12]. Demonstrating this lack of success for big data projects, Kelly and Kaskade [13] 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.

III. METHODS

A. Systematic Literature Review

A systematic literature review (SLR) is a way of identifying, evaluating and interpreting research relevant to a particular research question, topic area, or phenomenon of interest, using a research method that is reliable, accurate and facilitates auditing [14]. Our efforts to summarize existing research and to identify gaps in current research with respect to the process and methodologies teams use to execute Big Data projects followed the SLR guidelines outlined by Kitchenham [15].

B. Sources

The first step in doing an SLR is to describe the strategy used to conduct the search for published articles. In our case, to identify relevent articles, we used both manual searches of specific journals and conferences proceedings as well as automatic searches of electronic databases.

1) Manual Search of Selected Publications

With respect to the manual search of conference proceedings and journals, the SLR on which we report was conducted on the papers published in top tier conferences and journals within the information systems, data science, big data and business intelligence fields. Since Big Data is a new field, we focused our manual search on the two most recent years (2014 and 2015).

The conferences and journals were selected according to the four criteria: description (scope), subject area, subject category and SRJ ranking. Journals with a priori very technical content were excluded, e.g. IEEE Transactions on Information Theory.

In all, as shown in Table 1, six conferences and thirty-three journals were included in the manual part of our literature review.

Source	Name of Journal / Conference
Conference	IEEE Big Data
Conference	IEEE Big Data Congress
Conference	Data Science and Advanced Analytics
Conference	International Conference on Information Systems
Conference	Hawaii International Conference on System Sciences
Conference	Americas Conference on Information Systems
Journal	Journal of Information Technology
Journal	European Journal of Information Systems
Journal	Information Processing and Management, Scientometrics
Journal	International Journal of Information Management
Journal	Annual Review of Information Science and Technology
Journal	Information Technology and People, Decision Support Systems
Journal	Communication of the AIS
Journal	Journal of Management Information Systems
Journal	INFORMS Management Science
Journal	INFORMS Information Systems Research
Journal	Journal of the Association for Information Systems
Journal	INFORMS Journal on Computing
Journal	Journal of Operations Management
Journal	MIS Quarterly: Management Information Systems
Journal	Journal of the ACM
Journal	Enterprise Information Systems
Journal	Operations Research
Journal	Artificial Intelligence
Journal	Information Sciences
Journal	Research Journal of Information Technology
Journal	Surveys in Operations Research and Management Science
Journal	IEEE Transactions on Knowledge and Data Engineering
Journal	IEEE Transactions on Big Data
Journal	Big Data
Journal	Data Science Journal
Journal	EPJ Data Science
Journal	Journal of Big Data
Journal	Journal of Data Science
Journal	Journal of the Association for Information Science and Technology
Journal	Intelligence and Analytics
Journal	IEEE Intelligent Systems
Journal	International Journal of Technology Management
Journal	IEEE Internet Computing

TABLE 1. Sources for the Manual Literature Review

2) Online Search Strategy

In addition to the conferences and journals noted in Table 1, a Google Scholar search was employed to get additional relevant papers. Table 2 shows the five different Google Scholar searches that were used to identify additional papers. The search terms included both “big data” and “data science” as well as topic areas such as “process methodology”, “team coordination” and “project management”. Note that since the field of Big Data is changing rapidly, we focused our analysis on papers published after 2010.

Source	Search Terms
Google	"big data" +"process methodology"
Google	"data science" +"process methodology"
Google	"data science" +"team coordination"
Google	("data science" OR "big data") AND "team coordination"
Google	"data science" +"project management"

TABLE 2. Sources for the Online Literature Search

C. Selection Criteria

Three criteria were derived to select the papers that, taken together, could provide the current state of the published literature that addressed our research questions.

The first criterion was focused on if the paper discussed the process (or methodology) the team used to execute the project. In other words, was there a discussion on how the project team worked together and coordinated their tasks, or in general, collaborated on the project? This criterion was also useful to identify papers describing best practices making valuable suggestions to data science teams about how to run their Big Data projects. Some steps of the big data project might be very challenging and a special methodological approach might need to be provided and, therefore, the second criteria was focused on the lifecycle (or phases) of the project, in that we searched for papers that discussed one or more phases of the project (including items such as quality assurance and data cleaning). Papers falling into this category could provide valuable insight about the most challenging steps of the big data project such as data preparation, quality assurance, etc. We also identified if there was a description of a phase, perhaps pointing out a best practice (such as planning or time spent in a specific phase of the effort). Papers about algorithms and other technical aspects of the phases were not counted as papers falling into the second criteria. The third criterion was if a paper was focused on the maturity and/or the key factors that help drive the success of a big data project. Integrating knowledge from multiple studies about maturity levels and success factors could provide data scientists with a systematic view on what are the important factors that should be focused on to help ensure the successfully execution of a Big Data project, note that these factors might be based on the characteristics of the organization or the project.

In addition, we only included publications that were in English, and as previously noted, we explored papers published after 2010.

D. *Selection and Coding Process*

The selection and coding process was comprised of four steps, which were adapted from [14].

Step 1: An automatic search was performed, based on the search criteria defined in Table 2. The results were initially assessed by their title and abstract. The studies considered possibly relevant to the context of our research were selected for additional review.

Step 2: A manual search was performed in journals and conferences proceedings previously identified in Table 1. We focused our review on the journals and conferences published in 2014, 2015 and, where possible, 2016.

In terms of our process, first, for the more general conferences, a quick review of the conference determined which tracks might have relevant papers and all papers in those tracks were included in the review. For the big data focused conferences and our identified journals, all papers were included in our review.

Step 3: In this step, the papers identified in Step 1 and Step 2 were assessed via a coding methodology. The coding of the selected papers was a collaborative process as described by Saldana [16]. For each of the identified papers, two researchers coded each paper to determine if the paper discussed any of the criteria previously mentioned. The coders focused on the introduction and conclusion sections, but the coders could read other sections if deemed appropriate.

If there was a discrepancy, a third researcher made the final decision. Overall, the two reviewers agreed 96% of the time, which is above the minimum percentage of 85% suggested by Saldana [16]. This suggests that the process was robust and repeatable. The studies considered relevant were selected for a detailed review in step 4.

Step 4: For the papers that were determined to be relevant via the coding of process just discussed, additional detailed analysis was performed. This detailed analysis included identifying possible methodology best practices documented in each of the identified papers.

E. *Identified Papers*

In total, 10,028 papers were identified from our initial review of the targeted conferences, journals and Google Scholar search output (i.e., steps 1 & 2). Specifically, 1,645 papers were identified via our Google search, 3,651 were identified from conference proceedings and 4,732 papers were identified from the targeted journals. During step 3, the 10,028 papers were reviewed and 92 were identified as relevant (0.9% of the papers identified). An in-depth review of these 92 papers was then conducted (in step 4) and 42 papers were identified to be most relevant. As shown in Table 3, eighteen of the papers focused on the process the team used (or should use), seventeen of the papers discussed the project phase or lifecycle, and seven papers discussed maturity models (and/or critical success factors for doing a Big Data project). Note that for Google search Table 3 represents papers discovered by Google search only and does not show studies that were discovered in manual search.

Source	**Process**	**Phase**	**Maturity**
Conferences	12	16	4
Journals	0	0	1
Google	6	1	2
Total	18	17	7

TABLE 3. Summary of Papers Reviewed in Step 4

IV. FINDINGS

In this section, we summarize our analysis of the 42 papers identified during our review. Table 4 shows the papers identified, grouped by how we identified the paper and the focus of the paper (note that the numbers in the table represent the reference number for the paper).

	Conference	Journal	Search
Process & Phase	[2] [17] [24] [25] [28] [30] [31] [32]	--	--
Process	[1] [38] [46]	--	[19] [20]

	[47]		[21] [22] [50] [51]
Phase	[18] [26] [27] [29] [34] [35] [36] [40]	--	[52]
Maturity	[42] [44] [48] [49]	[43]	[39] [41]

Table 4: Identified Papers

A. *Recent Growth*

First, it is interesting to note that this topic is starting to become discussed more frequently. Figure 1 shows the number of papers returned from the Google Scholar search as well as papers identified via the manual search. As one can see, there is a growing interest in the methodological aspects of big data projects (note that 2016 was not shown in the chart, due to only obtaining a partial view of 2016).

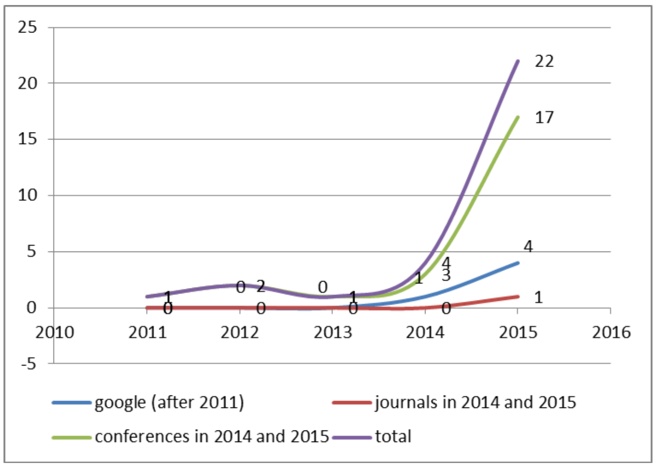

Figure 1: Trend of Relevant Articles

Taken together, this suggests a growing interest in this domain of study, and this trend is also born out by the fact that the first workshop on this topic was held in 2015[1].

B. *Team Process Insights*

The key insights, with respect to team process, are described below and include the need for a process, the key process attributes (such as iterative process model) needed within the process and the need for effective team communication.

1) *The need for a Process*

There is a growing recognition that projects need to focus on people, process and technology, not just analytics [17] [18] [19] and that a methodology would be useful [1]. However, it has been noted that two well-known Data Mining models, CRISP-DM (Cross-Industry Standard Process for Data Mining) and SEMMA (Sample, Explore, Modify, Model, Assess), might not be suitable for Big Data projects due its four V's [20].

In terms of the actual process methodology, it has been shown that Agile-like methodologies have several advantages over traditional waterfall-like methodologies [19] such as early time-to market, good collaboration, early risk revealing through iterative development [21]. Due to this, agile methodologies are starting to be discussed, such as an Agile-BI methodology [22]. However, in an experiment using multiple agile methodologies, it was shown that some agile methodologies might be better than other methodologies [23].

The goals of such a process methodology could be to improve coordination with others, ensure quality, ensure data ownership (as well as security and privacy), analyze requirements, prioritize requirements and enable analytical solutions to be deployed [2]. With this in mind, at least some managers are open to establishing a project methodology, but might not think of doing it unless prompted [24].

2) *Process attributes*

A critical success factor for executing Big Data projects is to use an iterative process model [17]. Saltz and Shamshurin [24], note that the process might be different for different phases of the project, perhaps with a more waterfall like data collection phase followed by a more iterative data analysis phase. In a similar finding, Vanauer [25] presents a waterfall-like workgroup ideation process, and then a phase to help determine the cost-benefit of the proposed project (i.e. the value of data and the data analytics). Furthermore, multiple process models might be needed and selecting the appropriate process model might be based on project attributes such as project clarity and complexity [1] or the type of project, such as a data driven or a question driven project [25]. Finally, it has been suggested that the methodology should include organizational capacity assessment and capacity building [26].

Another key suggestion is classify a project according to its difficulty (complex or simple) and clarity (clear or ambiguous) at the early stages of the project [1], which would allow different methodologies based on these characteristics. A domain dependent framework to carry out two types of projects (complex and ambiguous, simple and clear) was proposed in [1] that might be extended to other domains.

3) *The need for effective communication*

Effective communication is one of the key success factors for a Big Data project. For example, frequent dialog with senior management has been noted as being very important [24].

Gao, Koronios & Selle [17] note that an additional critical success factor for executing a big data project is to include multidisciplinary teams. Hence, a cross functional team might be needed to get maximum benefit from Big Data project [20]. This team might consist of people such as IT experts, scientists and business decision makers.

[1] http://www.midp.info/2015-workshop.html

However, multifunctional, often distributed, teams increase organizational complexity and increase the need for effective communication. In addition, collaboration with extended teams, such as IT departments, might require additional effort and may delay the overall project progress [27].

This suggests that a well-defined team process could be beneficial as it could drive proactive behavior and increase effective communication [1]. One advantage of a robust process is that it would typically improve group communication. Another way to improve communication is to implement a rotation program to increase the team's breadth of knowledge [1].

C. *Project Execution Insights*

The key insights, with respect to project execution, are grouped below by suggestions that could be used across all phases, and phase specific insights.

1) *Suggestions across the entire project*

Using a process based on CRISP-DM, Asamoah & Sharda [28] propose guidelines for each phase (ex. how to clean and validate data). While the guidelines are specific to one domain, these guidelines could likely be generalized. In addition, one could measure big data competence in each phase of the data life cycle (collection, store, process and disseminate), rather than just focus on the analytics phase [29].

In addition to these suggestions, some key challenges that have been noted include how to mitigate error propagation and data uncertainty, and more generally, how to measure accuracy [30]. Hence, there is a need for mechanisms to evaluate the validity of the results [31] (El-Gayer & Timsina, 2014) and ensure the analytical results are presented in an appropriate format for the end-user [31] (El-Gayer & Timsina, 2014).

2) *Phase specific insights*

Several efforts focused on a specific phase of the project. For example, the first step in a research project should be getting clarity about the purpose [32] (Das et al, 2015) and understanding the business objectives [33].

To structure and formalize business requirements for Big Data projects, Priebe and Markus [34] present an integrated methodology that links business concepts with data via a business information model.

In a different requirements suggestion, Di Tria [35] describes a data storage requirement phase that includes defining strategic goals, decision goals (to answer how strategic goals can be satisfied), and information goals (to answer decision goals). In addition, Corral [27] proposes a methodology for model formulation and retrieval, including a Model Management Warehouse that supports the methodology. In a related requirement analysis effort, a business requirement model, which is based on an extension of CRISP-DM, is suggested that includes determining business objectives, assessing the situation, determining the data mining goals and producing a project plan [33].

Other phases were also discussed. For example, a research phase has been suggested to understand how similar business problems were solved by other companies and scientists and to determine if a team's current IT infrastructure could support the data analysis methods required for the project [20]. With respect to analytical tools and models, an analytical governance framework could be useful. Such a framework could include data, IT, analytics as well as a results-sharing governance capability [36].

One noted benefit of defining specific phase standards is that a standard for data pre-processing could unify data gathering and integration, and a standard for data mining postprocessing could unify the models deployment [37]. Another benefit could be that a reference architecture, that leverages CRISP-DM for the analytical layer, could create a template for other organizations to use [38].

D. *Maturity Levels and Success Factors Insights*

Data science maturity describes how effective organizations employ their tools, people and other resources to manage and analyze data for the purpose of informing business decisions. Organizations with a high level of maturity run their projects holistically, with checks, feedback loops and mechanisms for improvement [39].

With respect to using maturity levels, Studer and Leimstoll [40] provide a step-by-step analytics process that takes into account the organization's analytics maturity across capabilities, culture and technology, in other words, they suggest that the steps executed should be a function of the organization's Big Data readiness and Booth [39] discusses some initial ideas with respect to maturity levels specific to Big Data. Other papers explored maturity models in related areas such as analytics and business intelligence [39] [41].

A related concept is that of critical success factors – what could an organization do to improve their chance of delivering a successful project. It has been noted that there is no single factor defining the success of the project, but rather, there are multiple such factors [42]. Others have also studied success factors [41] [43] [44] [17] [45] [1] .

Integrating these suggestions, several characteristics of a mature organization running Big Data projects were identified:

- Data (ability to store and access appropriate data)
- Governance (well defined roles and responsibilities)
- Process (using a formal methodology such as Agile)
- Objectives (with measurable KPIs)
- Team (skills in data-driven decision-making)
- Tools (to enable data-derived insights)

In addition, Table 5 shows an integrated view of these documented success factors, grouped by the previously identified characteristics of a mature organization. This table is extension of a CSF framework suggested by [45] in which three major groups of CSFs were identified: people, process and technology.

Data
• Data & data quality management / ownership
• Data Integration & Security
• Unstructured/structured data
• Representativeness of data
• Document collection/access to sources
Governance
• Management priority / sponsorship / support
• Big Data strategy alignment (with organization's vision)
• Project management process defined
• Well defined organizational structure
• Performance management
• Data protection and privacy by design
• Culture of being Data-driven
Process
• Close collaboration between IT and business
• Communication about the data and initiatives
• Flexibility and agility, with freedom for experimentation
• Focus on change management
• Project difficulty explored and communicated
• Clarity of project deliverables (clear or ambiguous)
Objectives
• Focus on small projects and known questions
• Specified business case
• Feasibility study
• Skill gap analysis
• Well defined scope – that understood by the team
• Measurable project outcome
Team
• Development of skills / training
• People skills & ability to self-organize when needed
• Data science, technology, business & management skills
• Multidisciplinary team (i.e., across different departments)
• Stakeholder coordination / shared understanding
Tools
• Investment in IT infrastructure, technology & tools
• Investment in data sources & data storage
• Reporting and visualization technology
• Discovery technology

TABLE 5. Success factors for Big Data Projects

V. CONCLUSION

This paper reports on a comprehensive review of the recent publications relating to the methodologies teams should use to execute Big Data projects. While there has not been significant research in this field, we note that the number of papers related to this domain has recently increased substantially, from seven in 2014 to twenty-three in 2015.

The contribution of the paper is in the following three outcomes. First, we identified the most valuable papers with respect to how to execute Big Data projects using a three-criterion based framework to identify Big Data methodology related papers. Second, we synthesized the knowledge from these studies. Our findings suggest that there is currently no agreed upon standard for executing these projects and that an improved process methodology would be useful. Our third contribution is an integrated set of success factors to help practitioners execute Big Data projects. Specifically, we defined thirty-three success factors, grouped by the six key project characteristics.

Finally, there were areas of research that did not show up in our review. For example, we saw no experiments comparing the effectiveness of how teams operate using different methodologies. In addition, while some case studies were identified, a clear next step could be additional case studies. Yet another next step could be to prioritize, refine, and validate our list of identified success factors. Finally, all CSFs can be ranked according to their importance for the project's success.

VI. REFERENCES

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

[2] 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 (Big Data).

[3] 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).

[4] 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)

[5] Shearer, C. 2000. "The CRISP-DM model: The new blueprint for data mining," Journal of Data Warehousing, 5(4).

[6] Matignon, R. 2007. "Data Mining Using SAS Enterprise Miner". John Wiley & Sons.

[7] Piatetsky, G. 2014. "CRISP-DM, still the top methodology for analytics, data mining, or data science projects," KDD News. DOI= http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

[8] Mohan, K., and Ahlemann, F. 2011. "What Methodology Attributes Are Critical for Potential Users? Understanding the Effect of Human Needs". Advanced I'lformation Systems Engineering: Springer, pp. 314-328.

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

[10] Chandler, N., & Oestreich, T. W. 2015. "Use analytic business processes to drive business performance". Retrieved from https://www.gartner.com/doc/2994617

[11] Naur, P., and Randell, B. 1969. “Software Engineering: Report of a Conference” Sponsored by the Nato Science Committee, Garmisch, Germany,7-11 Oct.1968, Brussels, Scientific Affairs Division, Nato.

[12] Goodwin, B. 2011. "Poor Communication to Blame for Business Intelligence Failure, Says Gartner." ComputerWeekly.com. Retrieved 24 February 2016, from http:/fwww.computerweekly.comfnews/1280094776/Poor-communication-to-blame-for-business- intelligence-failure-says-Gartner

[13] Kelly, J., & Kaskade, J. (2013). CIOS & BIG DATA What Your IT Team Wants You to Know. DOI= http://blog.infochimps.com/2013/01/24/cios-big-data/

[14] Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(TR/SE-0401), 28. http://doi.org/10.1.1.122.3308

[15] Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. Engineering, 45(4ve), 1051. http://doi.org/10.1145/1134285.1134500

[16] Saldana, J. 2008. Coding Manual for Qualitative Researchers, Sage Publications, Inc.

[17] 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), Puerto Rico

[18] 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 (Big Data).

[19] Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700–710.

[20] Dutta D, Bose I (2015) Managing a big data project: the case of Ramco Cements limited. Int J Prod Econ. doi:10.1016/j.ijpe.2014.12.032

[21] Dharmapal S., Sikamani T. (2016). Big Data Analytics Using Agile Model. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2016.

[22] A. A. R. E. Sheikh, Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications, 1.a ed. IGI Global, 2011.

[23] Saltz J., Crowston K., and Shamshurin I. 2017 Comparing Data Science Project Management Methodologies via a Controlled Experiment. Hawaii International Conference on System Sciences, (in press).

[24] 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 (Big Data).

[25] Vanauer M., Böhle C., Hellingrath B. 2015. “Guiding the Introduction of Big Data in Organizations: A Methodology with Business-and Data-Driven Ideation and Enterprise Architecture Management-Based Implementation”, Hawaii International Conference on System Science (HICSS)

[26] Kee K. 2015. “Three Critical Matters in Big Data Projects for e-Science”, IEEE International Conference on Big Data (Big Data).

[27] Corral, K., Schuff D., Schymik G., Louis R. 2015. “Enabling Self-Service BI through a Dimensional Model Management Warehouse”, Twenty-first Americas Conference on Information Systems (AMCIS).

[28] Asamoah D., and Sharda R. 2015. “Adapting CRISP-DM Process for Social Network Analytics: Application to Healthcare”, Twenty-first Americas Conference on Information Systems (AMCIS)

[29] Kung L., Kung H., Jones-Farmer A., Wang Y. 2015. “Managing Big Data for Firm Performance: a Configurational Approach”, Twenty-first Americas Conference on Information Systems (AMCIS)

[30] Dorr B., Greenberg C., Fontana P., Przybocki M., Bras M., Ploehn C., Aulov O., Michel M., Golden E., Chang W. 2015. “The NIST Data Science Program”, International Conference on Data Science and Advanced Analytics (DSAA), Paris, France

[31] El-Gayar O., Timsina P. 2014. “Opportunities for Business Intelligence and Big Data Analytics in Evidence Based Medicine”, Hawaii International Conference on System Science (HICSS)

[32] 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 (Big Data).

[33] Niño M, Blanco JM, Illarramendi A. “Business Understanding, Challenges and Issues of Big Data Analytics for the Servitization of a Capital Equipment Manufacturer”. Proceedings of the 2015 IEEE International Conference on Big Data. October 2015, p.1368-1377. DOI: http://dx.doi.org/10.1109/BigData.2015.7363897

[34] Priebe T., Markus S. 2015. “Business Information Modeling: A Methodology for Data-Intensive Projects, Data Science and Big Data Governance”, IEEE International Conference on Big Data

[35] Di Tria F., Lefons E., Tangorra F. 2014. “Design process for Big Data Warehouses”, International Conference on Data Science and Advanced Analytics (DSAA), Paris, France

[36] Avery A., Cheek K. 2015. “Analytics Governance: Towards a Definition and Framework”, Twenty-first Americas Conference on Information Systems (AMCIS), Puerto Rico

[37] Esfandiari, N., Babavalian, M. R., Moghadam, A.-M. E., Tabar, V. K. (2014). Knowledge discovery in medicine: Current issue and future trend. Expert Systems with Applications, 41(9), 4434-4463.

[38] Siriweera T.H.A.S., Paik I., Kumara B., Koswatta K.R.C. 2015. “Intelligent Big Data Analysis Architecture Based on Automatic Service Composition”, IEEE International Congress on Big Data

[39] Booth D. (2015). Digital Success through Data and Analytics Maturity. Journal of Digital & Social Media Marketing, vol. 2 (4), Feb. 2015, pp. 338-344(7)

[40] Studer Y., Leimstoll U. 2015. “Improving Consulting Processes in Web Analytics: A Framework for Multichannel Analytics”, Twenty-first Americas Conference on Information Systems (AMCIS).

[41] Muller L., Hart M. (2016). Updating Business Intelligence and Analytics Maturity Models for New Developments. 2nd International Conference, ICDSST 2016, Plymouth, UK, May 23–25, 2016, Proceedings, pp. 137-151.

[42] Cato P, Gölzer P, Demmelhuber W (2015) An investigation into the implementation factors affecting the success of big data systems. In: 11th Int Conf Innov Inf Technol. S 134–139

[43] Brooks, P., El-Gayar, O., and Sarnikar, S. (2015) A Framework for Developing a Domain Specific Business Intelligence Maturity Model: Application to Healthcare. International Journal of Information Management, 35(3): 337–45.

[44] Cosic, R., Shanks, G., & Maynard, S. (2012). Towards a business analytics capability maturity model. 23rd Australasian Conference on Information Systems, Geelong.

[45] Sicular, S. (2012). No Data Scientist Is an Island in the Ocean of Big Data. Gartner.

[46] Guo P. (2013) Data Science Workflow: Overview and Challenges, CACM blog

[47] Ahangama, S., and Poo, C. C. D. 2015. “Designing a Process Model for Health Analytic Projects,” PACIS 2015 Proceedings. Paper 3.

[48] Sathi, A. (2012) Big Data Analytics: Disruptive Technologies for Changing the Game, Mc Press

[49] K Crowston & J Qin (2011). ‘A Capability Maturity Model for Scientific Data Management: Evidence from the Literature’. In: Proceedings of the American Society for Information Science and Technology. Vol. 48. doi: 10.1002/meet.2011.14504801036.

[50] Ahangama, S., Poo, D.C.C.: Unified structured process for health analytics. Int. J. Med. Health Pharm. Biomed. Eng. 8, 744–752 (2014)

[51] O. Niakšu, “CRISP Data Mining Methodology Extension for Medical Domain,” Balt. J. Mod. Comput., vol. 3, no. 2, pp. 92–109, 2015.

[52] Niño M., Blanco J., Illarramendi A., Business understanding, challenges and issues of Big Data Analytics for the servitization of a capital equipment manufacturer, 2015 IEEE International Conference on Big Data, pp. 1368-1377, DOI: 10.1109/BigData.2015.7363897