


Data Analysis for Software Process Improvement: A Systematic Literature Review

Jezreel Mejía , Freddy Íñiguez, and Mirna Muñoz

Center for Research in Mathematics, Avenida Universidad 222, 98068 Zacatecas, Mexico
{jmejia, freddy.iniguez, mirna.munoz}@cimat.mx

Abstract. Current information systems demand high quality software products that guarantee a safety and a reliable use for our day-to-day life. A common understanding between software organizations and practitioners is that software product quality largely depends on the software process quality. A Software Process Improvement (SPI) initiative consists of a set of practices and activities that are designed to improve software organizations processes through the evaluation of their current practices and the way software products and services are developed. However, the big amount of information that is generated from the software organization practices has complicated the knowledge extraction, and therefore, the SPI initiatives. A possible technique to make a good knowledge management is data analysis. This paper presents the results of a systematic literature review to establish the state-of-the-art of data analysis for software process improvement. The findings also encourage to the creation of a BigData-based data analysis model in a future work for this research.

Keywords: Knowledge management · Information management · Data analysis · Software product quality · Software process quality · Software engineering

1 Introduction

Information systems have become an essential part of our lives. Nowadays, one can use a system or application in practically any discipline, like in medicine, economics, architecture, education, astronomy, psychology, law, and mathematics. Such demands for information systems have made software development organizations to create software products and services of high quality, that guarantee a safety and a reliable use, and focused to a specific purpose [1].

There is a common understanding between software development organizations and practitioners, which is that software product quality largely depends on the software process quality itself [2].

A Software Process Improvement (SPI) initiative is a set of practices and activities that are designed to improve the software organizations processes through the evaluation of the current practices and the way software products and services are developed, considering the experiences and competencies of the organization's practitioners [3]. In addition, Knowledge Management (KM) is a discipline that take advantage of this

information to provide innovation, responsiveness, competency, and efficiency, making KM an essential practice to be adopted by software organizations to succeed in SPI [3].

Moreover, SPI initiatives have become an essential tool for project managers and software engineers to achieve their goals, maintain their competitiveness, and to provide a return on investment to the organizations [4].

However, the big amount of information from practices and activities during the development of a software product has complicated the extraction and management of knowledge, and therefore, SPI initiatives. According to [5], since this increment of information, a possible technique to help practitioners in the detection of implicit and useful information with lower cost, high correctness, high noise filtering rate, and fit for large databases is data analysis.

The purpose of this research study is to perform a Systematic Literature Review (SLR) to obtain the most significant amount of recent studies related to the way data analysis techniques are applied in software development organizations with the main purpose of understanding how knowledge is managed to SPI.

This paper is organized as follows: Section 2 details the phases and steps indicated by the SLR. Section 3 presents the results obtained after the analysis of the primary studies, as well as some discussions about the findings, while Sect. 4 covers conclusions and future work.

2 Research Methodology

A Systematic Literature Review (SLR) is a mean which helps to identify, evaluate, and interpret all available studies around a specific research question. This protocol is presented by [6] and expresses how to perform a SLR by following three main phases: (1) Planning the review, (2) Conducting the review, and (3) Reporting the review. These three phases are developed and explained in the following sub-sections.

2.1 Planning the Review

This phase is focused on establishing the need to perform the SLR, in addition to the definition of the research questions that guide the entire process of the SLR and the selection of the data sources and scientific databases to extract the primary studies.

2.1.1 Identify the Need to Perform the SLR

Given the current problematic mentioned in the introduction section about the increment of information in software organizations, it becomes necessary to know the existing approaches in the application of data analysis techniques for SPI. In the same way, a well understanding of the main problematics and challenges of both disciplines, data analysis and SPI, will help to develop a data analysis model which integrates a solution for the current data and information requirements.

2.1.2 Define the Research Questions

The Research Questions (RQ) for this study are as follow:

- RQ1.* Is there any study on the use of data analysis for SPI?
- RQ2.* What are the existing approaches related to data analysis for business or decision-making processes?
- RQ3.* What are the main data analysis techniques applied to SPI?
- RQ4.* What are the main problematics and challenges in the application of data analysis techniques for SPI?
- RQ5.* What are the main problematics and challenges about the way organizations are managing the knowledge/data generated from their processes?

2.1.3 Create the Search String

In order to facilitate the creation of the search string, a set of keywords have been identified from the research questions in the previous step, which are the following: (1) *data analysis*, (2) *software process improvement*, (3) *decision-making process*, and (4) *business process*.

Through a keywords combination and the use of logical connectors “AND” and “OR”, the resulting search string is as follows:

“(Data Analysis AND ((Software Process Improvement) OR ((Decision-making) OR (Decision-making Processes)) OR (Business Process) OR (Industrial Process)))”

Table 1. Inclusion and Exclusion criteria.

Inclusion criteria	Exclusion criteria
Title and keywords sections contain at least two of the keywords	Title does not contain any of the keywords
Abstract is related to data analysis for SPI	Studies not related to data analysis or SPI
Studies containing case studies and practical applications of data analysis at industrial or academic fields	Studies related to SPI, but not to data analysis
Studies containing approaches of data analysis methodologies, methods, or techniques for SPI	Studies related to data analysis, but also to other fields like health, economics, politics, or agriculture
Studies containing results of the application of data analysis for SPI	Studies not in the English or Spanish language
Studies related to the decision-making process based on the knowledge and data of the organizations	Duplicated studies; in this case the most recent study will be kept
Studies in the English or Spanish language	Inaccessible studies
Studies published between 2010 and 2016	Studies from before 2010

2.1.4 Select the Data Sources

The following data sources were used to the extraction of studies, mainly due to their highly use and reference in both the scientific and software engineering fields: (a) *IEEE Explore*, (b) *ACM Digital Library*, (c) *Springer Link*, and (d) *Web of Science*.

2.2 Conducting the Review

The second phase of the SLR, the definition of the inclusion and exclusion criteria are performed, in addition to the selection of the primary studies through the definition of a study selection process and a further application of some study quality assessment.

2.2.1 Set the Inclusion and Exclusion Criteria

In order to limit the number of results and avoid unnecessary or not-relevant studies for this research, it is critical to define some inclusion and exclusion criteria. These criteria are given in Table 1.

2.2.2 Select the Primary Studies

During this step, the SLR protocol [6] suggests to define a study selection process to obtain the primary studies for this research. This process consists of four steps: (1) apply and adapt the search string to every research library and scientific database, (2) filter studies by reading the title and apply the first inclusion criteria, (3) apply the rest of the inclusion and exclusion criteria on reading abstracts, introductions, conclusions, and if necessary, all the study, and (4) select the primary studies.

Figure 1 shows how at the end of this study selection process, the 25 studies from the third step were selected and used for this research:

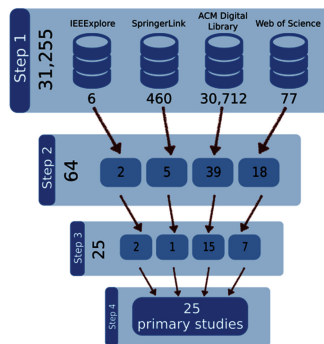


Fig. 1. Study selection and results

2.2.3 Study Quality Assessment

The study quality assessments guarantee that the information contained in each of the primary studies is relevant and valuable for this research, avoiding an extra effort when

analyzing the study in detail and saving time. The following list represents the Study Quality Assessments:

SQA1: Does the study focus on the application and use of data analysis techniques for software process improvement?

SQA2: Does the study include the main implications and challenges related to the application of certain data analysis techniques?

SQA3: Does the study explain how to use the knowledge and data of an organization to improve their processes?

The primary studies were evaluated with these previous study quality assessments and it has been decided to keep all the 25 primary studies after this evaluation.

3 Results

This section presents the SLR results from 25 primary studies.

3.1 Data Analysis Approaches for SPI

Given *RQ1*, about the studies that use data analysis for SPI, the SLR have highlighted the 25 primary studies, which include some interesting approaches that have been applied in a variety of fields for the recent years, specifically in the fields of software engineering [7–9], manufacturing [10], and business intelligence for Small and Middle-Size Enterprises (SME) or novice practitioners [11, 12].

Regarding to *RQ2*, from the 25 primary studies, 17 of them are focused in the application of data analysis for software processes. Moreover, 6 of the studies are related to the application of data analysis for a business process orientation, meantime just 2 of the studies are focused for an industrial process.

3.2 Data Analysis Techniques for SPI

According to *RQ3* related to the main data analysis techniques that are used in the SPI field, a variety of approaches have been obtained, where each of the approaches make use of different parameters and information to obtain knowledge and improve the business. From a general perspective, these data analysis techniques are primarily focused in obtaining valuable information from software repositories [13], defects/bug databases [8], peer-code review systems, mailing lists, online forums [14], production data, and even from reports and postmortems meetings [15]. Thus, this information is analyzed and finally presented to project managers and developers in a useful way which help them to improve their decision-making.

Mining Software Repositories (MSR for short) is the most used technique found in the primary studies followed by a combination of different approaches, which are considered by some authors as *classical* data analysis techniques, this because their wide use and adaptation. These approaches are regression, classification, and clustering.

Likewise, there are other approaches that were highlighted by the SLR, like Orthogonal Defect Classification, Computer Cognitive Resonance, Multivariate analysis, Anomaly detection, Common Warehouse Metamodel, Frequent Sequence Mining, and Home-grown matching.

3.3 Issues and Challenges of Data Analysis

Through the analysis of the primary studies, it is possible to identify the main issues, concerns, and challenges around the use of data analysis for the SPI. These following findings are related to the initial concerns of *RQ4*:

- *Data is not the problem.* Today's software development organizations generate enough data from their own practices in the form of event logs, software repositories, bug reports, commit history, test suits, documentation, among others [13, 16–18]. However, from all this amount of information, very little is organized, cleansed, standardized, and most importantly, presented in a way that is useful to developers and project managers to support their decision-making. Thus, practitioners make daily decisions based on their previous experiences, consulting other developers, or just by their intuition and good-feeling.
- *Exponential growth of data.* What it is a problem is the steadily increase of data and information in a software, business, or industrial organization. The big amount of data generated during the development process or the production process it becomes unable to process for the most of the companies. It is therefore that some researches [16, 17, 19] are focused in developing a capable infrastructure to deal with the big data needs nowadays.
- *Expertise required to analyze data.* Today's novice data scientists experiment a lack on the skills that are necessary to extract the information from the data resources, to correctly select the data analysis techniques, and to accurately interpret the results. This is an important aspect, given that just the correct selection of attributes and algorithms have an impact in the success of the data analysis [12]. Here comes the importance of some researches which efforts are focused in develop adapted frameworks for novice practitioners or SME, like in [11, 12].
- *Data details.* The level of data details comprises a significant aspect. For example, in [8] they implemented an Orthogonal Defect Classification technique to support the root cause analysis for a software enterprise, and to have a successful root cause analysis these defects logs should be briefer and detailed. Here is where the proposal of [20] becomes essential, given that its main purpose is to create a novel specialized compression technique for process logs, which guarantee no loss of information.
- *Understanding of data.* To be successful in a data analysis project efforts must come from most stakeholders, i.e. from project managers and directives to software developers, data scientists, and even customers. Thus, experts suggest a deep analysis of the data and information to have a better understanding about the data and its potential from different perspectives [18].
- *Presenting data.* The way information is presented is almost as essential as communicating the value of the results [18]. Thus, presentation of information and insights

after data analysis must be simple, intuitive, reliable, replicable, and should reflect the view of the business customer. Bertini & Lalanne [21] made an interesting research about the various ways information visualizations (commonly named just *in-fovis*) and data mining can be integrated to achieve effective knowledge discovery.

3.4 Issues and Challenges of Software Process Improvement

In addition to the implications and challenges related to data analysis in SPI, a second search was performed over *RQ5* to obtain the current issues and concerns around the way software organizations are managing the knowledge generated from the software development process itself. This second search was necessary to not interfere with the results of the first search (which comprises results for RQ1-RQ4). The results from second search come from the following search string:

“(Process) AND ((Information AND Management) OR (Knowledge AND Management))”

The search string above was performed on the same selected scientific databases in Sect. 2, obtaining the following results: (1) ACM: 12 699 articles, (2) Web of Science: 250 articles, (3) Springer Link: 11 071 articles, and (4) IEEE Explore: 3916. From a total of 27 936 articles, 8 of them were selected to conclude the following implications and challenges about the way software organizations are managing their process-generated information and knowledge:

- *Product quality affected by the software process quality.* The most of the studies obtained agree on the importance of SPI efforts to improve the quality of the software products and to maintain the competitiveness [2, 22, 23]. Chugh [3] says that the quality of the developed software is directly dependent on the quality of the development process, making a special interest in the adoption of standards like CMM, IDEAL, ISO 9001, among others to achieve the desired quality level.
- *Appropriate knowledge management process.* Several studies have emphasized the usefulness and benefits of an accurate Knowledge Management (KM) implementation in software organizations [3, 23, 24]. In addition, this KM strategy should come with an appropriate knowledge sharing activity to prevent the same mistakes, reduce dependency on those employees that own critical information, increase the integration of individual competencies, and improving the decision-making process [25].
- *Misunderstanding between information and knowledge.* A misunderstanding between information and knowledge exists, which complicates project managers and developers to take a full advantage of the valuable knowledge that exist on information and data. A simple but very interesting way to identify the differences between information and knowledge is explained in [26].
- *Rapid increase of data.* Given the advances in information technologies, the amount of information produced nowadays has increased exponentially, which occasionally makes difficult the knowledge extraction. According to [5], a possible solution to help practitioners in the detection of implicit and useful information with lower cost, high correctness, and fit for large databases is data mining.
- *Need for technical solutions.* Recent researches have noticed the need of technical solutions that enable flexible storage, retrieval, processing, and interpreting

information. Such is the case of [25], which emphasize the need of appropriate tools to support the knowledge characteristics.

3.5 Discussions About the Findings

The following list encompasses some concerns that came up once all the primary and secondary studies' issues and challenges have been analyzed.

- *Critical process areas.* SPI initiatives are designed to improve the software process. However, the most of the studies did not mention why they selected specific areas of the business to improve. Thus, some concerns about this aspect are: What are the most critical areas to improve in a software development organization? Which areas provide the better Return on Investment (ROI)?
- *Business objectives and perspectives.* SPI initiatives consider the business objectives and perspective, even though a SPI initiative is designed to improve the process. Why does the data analysis strategy should reflect the view and expectations of the business customer?
- *Data selection and extraction.* Related to the selection of the area to improve in the first discussion point, just a few studies explicitly justified their selection of a specific kind of data. However, how can we identify the kind of data that is going to be useful for the data analysis process? Do software organizations already own the accurate information or it is necessary to generate/obtain it?
- *Data cleansing.* None of the studies mention any related problematic about the cleansing and preparation of the raw data to analyze, even though this is quite a challenge in the data analysis process [27]. Thus, how can we clean the data? Is all the data necessary? What are the most recommended tools to perform a cleansing of data? Is it possible to clean any kind of data?
- *Understanding the data.* In a hypothetical situation where one already have selected and cleansed the data, why is it important to get familiarized with and to understand the data? Having a well understanding about the information, does it affect any part of the data analysis process? What are the benefits?
- *Data analysis technique selection.* A vital step to guarantee the success of a data analysis effort is the selection of the data analysis technique to be applied. However, the most of the studies did not justify the application of a specific data analysis technique. Being such an important step, what is the best guide to the selection of the techniques or approaches for the data analysis process? What aspects are involved in the selection of such techniques? Is it possible to customize or change any data analysis technique during the process?
- *Suggest the improvements.* Regardless the existence of some interesting approaches about the presentation of the results and improvements, the most of the studies did not indicate the ways to present the results or at least give an orientation about how to make the suggestions for SPI. Is there any standard or guide to process the results and orient them to SPI? Is there a graph for a specific kind of data? Are there specific views for specific stakeholders?

4 Conclusions and Future Work

A Systematic Literature Review (SLR) was performed to establish the state-of-the-art of data analysis for Software Process Improvement (SPI), focused on the current studies that apply data analysis in software, business, or industrial environments, in addition, to the implications and challenges of both fields, data analysis and SPI. From a total set of 31 255 studies, just 25 of them were selected as the primary studies for this research. The most relevant results indicate the increasingly demand on the application of data analysis techniques to facilitate the management of information and knowledge to improve the software processes, and thus increment the quality of the software products, maintain the competitiveness, and to improve relationships with customers.

Other results indicate the most-used data analysis techniques, like MSR and *classical* approaches that have been applied in a variety of environments like software, business, and industrial organizations.

Finally, it was possible to identify a set of implications and challenges around the use of data analysis for SPI. Thus, it was also possible to detect some deficiencies in the current approaches and efforts in data analysis for SPI, like a need to make a correct selection of the process areas to improve, to integrate the business objectives and perspective in the data analysis strategy, as well as an accurate selection of the data sources, data analysis techniques, and visualization tools to represent the results.

Based on these findings, a motivation source has been set for a further step in this research, which encourages the creation of a data analysis model that integrates a solution for the above-mentioned deficiencies. This data analysis model should consider the following aspects: (a) to be a Big Data approach to deal with the recent issues of the rapid increase of data and information, (b) it should help practitioners to identify the need for performing a data analysis, (c) it must consider the business objectives and perspectives to be integrated in the data analysis strategy, (d) it should help practitioners to make an accurate selection of the data sources and data techniques, and (e) it should consider an appropriate guide to present the results in an useful way to the most of the stakeholders.

Appendix 1: List of Primary Studies

[PS-01] Baysal, O. (2013). Informing development decisions: From data to information. International Conference on Software Engineering. <https://doi.org/10.1109/ICSE.2013.6606729>

[PS-02] Pavon, R., & Carpenter, B. (2013). Synthesis of Decision Making: From Data to Business Execution. <https://doi.org/10.1109/ICDMW.2013.42>

[PS-03] Ivarsson, M., & Gorschek, T. (2012). Tool support for disseminating and improving development practices. Software Quality Journal. <https://doi.org/10.1007/s11219-011-9139-6>

[PS-04] Leida, M., Majeed, B., Colombo, M., & Chu, A. (2013). LNBIP 162 - A Lightweight RDF Data Model for Business Process Analysis.

- [PS-05] Shibata, T., & Kurachi, Y. (2015). Big Data Analysis Solutions for Driving Innovation in On-site Decision Making, 51(2), 33–41.
- [PS-06] Roedder, N., Karaenke, P., Knapper, R., & Weinhardt, C. (2014). Decision-making based on incident data analysis. 16th IEEE Conference on Business Informatics, CBI 2014. <https://doi.org/10.1109/CBI.2014.47>
- [PS-07] Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2015). Real-time business activity monitoring and analysis of process performance on big-data domains. Telematics and Informatics. <https://doi.org/10.1016/j.tele.2015.12.005>
- [PS-08] Ogiela, L., & Ogiela, M. (2015). Semantic Data Analysis Algorithms Supporting Decision-making Processes. <https://doi.org/10.1109/BWCCA.2015.108>
- [PS-09] Esaki, K., Ichinose, Y., & Yamada, S. (2012). Statistical Analysis of Process Monitoring Data for Software Process Improvement and Its Application. American Journal of Operations Research, 2, 43–50. <https://doi.org/10.4236/ajor.2012.21005>
- [PS-10] Söylemez, M., & Tarhan, A. (2013). Using process enactment data analysis to support orthogonal defect classification for software process improvement. Joint Conference of the 23rd IWSM-MENSURA 2013. <https://doi.org/10.1109/IWSM-Mensura.2013.27>
- [PS-11] Fazzinga, B., Flesca, S., Furfaro, F., Masciari, E., & Pontieri, L. (2015). A compression-based framework for the efficient analysis of business process logs. <https://doi.org/10.1145/2791347.2791351>
- [PS-12] Begoli, E. (2012). A Short Survey on the State of the Art in Architectures and Platforms for Large Scale Data Analysis and Knowledge Discovery from Data.
- [PS-13] Serban, F. (2013). A Survey of Intelligent Assistants for Data Analysis. ACM Comput. Surv, 45(35). <https://doi.org/10.1145/2480741.2480748>
- [PS-14] Sureka, A., Kumar, A., & Gupta, S. (2015). Ahaan: Software Process Intelligence: Mining Software Process Data for Extracting Actionable Information. <https://doi.org/10.1145/2723742.2723763>
- [PS-15] Zheng, L., Zeng, C., Li, L., Jiang, Y., Xue, W., Li, J., Wang, P. (2014). Applying Data Mining Techniques to Address Critical Process Optimization Needs in Advanced Manufacturing. <https://doi.org/10.1145/2623330.2623347>
- [PS-16] Santos, T. A., Lima, A. M., Lima Reis, C. A., & Quites Reis, R. (2014). Automated Support for Human Resource Allocation in Software Process by Cluster Analysis. <https://doi.org/10.1145/2593822.2593830>
- [PS-17] Grabova, O., Darmont, J., Chauchat, J.-H., & Zolotaryova, I. (2010). Business Intelligence for Small and Middle-Sized Enterprises.
- [PS-18] Menzies, T., Kocaguneli, E., Peters, F., Turhan, B., & Minku, L. L. (2013). Data Science for Software Engineering.
- [PS-19] Bertini, E., & Lalanne, D. (2010). Investigating and Reflecting on the Integration of Automatic Data Analysis and Visualization in Knowledge Discovery.
- [PS-20] Houston, D. X., & Buettner, D. J. (2013). Modeling User Story Completion of an Agile Software Process.
- [PS-21] Mazón, J.-N., Zubcoff, J. J., Garrigós, I., Espinosa, R., & Rodríguez, R. (2012). Open Business Intelligence: on the importance of data quality awareness in user-friendly data mining.

- [PS-22] Rao, J. J., Kelappan, R., & Pallath, P. (2014). Recommendation System to Enhance Planning of Software Development using R. <https://doi.org/10.1145/2593822.2593831>
- [PS-23] Deline, R. (2015). Research Opportunities for the Big Data Era of Software Engineering. In Proceedings - 1st International Workshop on Big Data Software Engineering, BIGDSE 2015. <https://doi.org/10.1109/BIGDSE.2015.13>
- [PS-24] Zhang, D., Dang, Y., Lou, J.-G., Han, S., Zhang, H., & Xie, T. (2011). Software Analytics as a Learning Case in Practice: Approaches and Experiences.
- [PS-25] Marcus, A., & Menzies, T. (2010). Software is Data Too.

References

1. O'Regan, J.: Introduction to software process improvement. *J. Chem. Inf. Model.* **53**(9), 1689–1699 (2013)
2. Mejía, J., Muñoz, E., Muñoz, M.: Reinforcing the applicability of multi-model environments for software process improvement using knowledge management. *Sci. Comput. Program.* **121**, 3–15 (2016)
3. Chugh, M., Chugh, N., Punia, D. K.: Evaluation and analysis of knowledge management best practices in software process improvement a multicase experience. In: Second International Conference on Advances in Computing and Communication Engineering (ICACCE), pp. 661–666 (2015)
4. Kuhrmann, M., Konopka, C., Nellemann, P., Diebold, P., Münch, J.: Software process improvement: where is the evidence? initial findings from a systematic mapping study. In: Proceedings of the 2015 International Conference on Software and System Process, pp. 107–116 (2015)
5. Tsai, C.F., Yeh, H.F., Chang, J.F., Liu, N.H.: PHD: an efficient data clustering scheme using partition space technique for knowledge discovery in large databases. *Appl. Intell.* **33**(1), 39–53 (2010)
6. Kitchenham, B.: Systematic reviews. In: 10th International Symposium on Software Metrics (2004)
7. DeLine, R.: Research opportunities for the big data era of software engineering. In: 1st International Workshop on Big Data Software Engineering, BIGDSE (2015)
8. Söylemez, M., Tarhan, A.: Using process enactment data analysis to support orthogonal defect classification for software process improvement. In: Joint Conference of the 23rd IWSM-MENSURA (2013)
9. Rao, J., Kelappan, R., & Pallath, P.: Recommendation system to enhance planning of software development using R (2014)
10. Zheng, L., Zeng, C., Li, L., Jiang, Y., Xue, W., Li, J., Wang, P.: Applying data mining techniques to address critical process optimization needs in advanced manufacturing (2014)
11. Grabova, O., Darmont, J., Chauchat, J., Zolotaryova, I.: Business intelligence for small and middle-sized enterprises (2010)
12. Mazón, J., Zubcoff, J., Garrigós, I., Espinosa, R., Rodríguez, R.: Open business intelligence: on the importance of data quality awareness in user-friendly data mining (2012)
13. Baysal, O.: Informing development decisions: From data to information. In: International Conference on Software Engineering (2013)
14. Sureka, A., Kumar, A., Gupta, S.: Ahaan: Software process intelligence: mining software process data for extracting actionable information (2015)

15. Ivarsson, M., Gorschek, T.: Tool support for disseminating and improving development practices. *Softw. Qual. J.* **20**, 173–199 (2012)
16. Shibata, T., Kurachi, Y.: Big data analysis solutions for driving innovation in on-site decision making. *Fujitsu Sci. Technol. J.* **51**(2), 33–41 (2015)
17. Vera, A., Colomo, R., Molloy, O.: Real-time business activity monitoring and analysis of process performance on big-data domains. *Telematics Inform.* (2015)
18. Pavon, R., Carpenter, B.: *Synthesis of decision making: from data to business execution* (2013)
19. Leida, M., Majeed, B., Colombo, M., Chu, A.: A lightweight RDF data model for business process analysis. In: Cudre-Mauroux, P., Ceravolo, P., Gašević, D. (eds.) *SIMPDA 2012. LNBP*, vol. 162, pp. 1–23. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40919-6_1](https://doi.org/10.1007/978-3-642-40919-6_1)
20. Fazzinga, B., Flesca, S., Furfaro, F., Masciari, E., Pontieri, L.: A compression-based framework for the efficient analysis of business process logs (2015)
21. Bertini, E., Lalanne, D.: *Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery* (2010)
22. Chang, C., Lin, T.: The role of organizational culture in the knowledge management process. *J. Knowl. Manage.* **19**(3), 433–455 (2015)
23. Diedrich, A., Guzman, G.: From implementation to appropriation: understanding knowledge management system development and introduction as a process of translation. *J. Knowl. Manage.* **19**(6), 1273–1294 (2015)
24. Balco, P., Drahoova, M.: Knowledge management as a service (KMaaS). In: *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 57–62 (2016)
25. Lee, K., Chen, Y., Muñoz, C.: Examining the impact of organizational culture and top management support of knowledge sharing on the success of software process improvement. *Comput. Hum. Behav.* **54**, 462–474 (2016)
26. Lihua, L., Feifei, Y.: Knowledge management in high technology enterprises. In: *2010 International Conference on E-Business and E-Government*, pp. 1823–1826 (2010)
27. Cuesta, H.: *Practical Data Analysis*. Packt Publishing, Birmingham (2013)