



On Applicability of Big Data Analytics in the Closed-Loop Product Lifecycle: Integration of CRISP-DM Standard

Elaheh Gholamzadeh Nabati, Klaus-Dieter Thoben

► To cite this version:

Elaheh Gholamzadeh Nabati, Klaus-Dieter Thoben. On Applicability of Big Data Analytics in the Closed-Loop Product Lifecycle: Integration of CRISP-DM Standard. 13th IFIP International Conference on Product Lifecycle Management (PLM), Jul 2016, Columbia, SC, United States. pp.457-467, 10.1007/978-3-319-54660-5_41 . hal-01699721

HAL Id: hal-01699721

<https://inria.hal.science/hal-01699721>

Submitted on 2 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

On Applicability of Big Data Analytics in the Closed-Loop Product Lifecycle: Integration of CRISP-DM Standard

Elaheh Gholamzadeh Nabati¹, and Klaus-Dieter Thoben²,

¹International Graduate School for Dynamics in Logistics, University of Bremen, Bremen, Germany

nab@biba.uni-bremen.de

²BIBA - Bremer Institute für Produktion und Logistik GmbH, Bremen, Germany

tho@biba.uni-bremen.de

Abstract. The product use data can have an important role in closed-loop product lifecycle management (CL-PLM), where information feedbacks from the use data can contribute to improve the product design and performance. The product usage data can nowadays be collected easier than before, with the aid of sensors and technologies embedded in products. However, the collected data can have complex characteristics. They come from various sources, have different formats and high volume. In order to improve the product lifecycle processes with these data, discussing the use of data analysis in the product lifecycle is necessary. Analyzing the data with such characteristics has been also considered in the context of big data analytics. In this paper an approach for standardization of the process of usage data analysis based on a standard called Cross Industry Standard Process for Data Mining (CRISP-DM), is introduced and its potential integration in CL-PLM is investigated. The reference steps of analyzing usage data are identified. They cover the processes between data generation until feeding back the knowledge of use to the product design phase.

Keywords: Big data, Closed-loop product lifecycle, CRISP-DM, Product use data, Standardization

1 Introduction and Problem Description

Product lifecycle is referred to the different stages which product goes through during its life span. Usually these steps are seen from the time the concept of the product is developed until the retirement and recycling. These stages can be grouped into three

phases. Beginning of the life (BOL), middle of the life (MOL) and end of the life (EOL). BOL includes product development and manufacturing. MOL starts when the product is put into operation and EOL refers to the activities of recycling and disposing [24], [27]. Closed-loop PLM focuses on enabling continuous flow of information between different phases of the lifecycle [10]. Improving the information flows includes better use of product lifecycle data and enhancing the processes of the lifecycle with them. For example, improving the product development and design process with the product use data. In this regard, accurate understanding of product operating condition helps design engineers to find the cause of failures. Thus, it can increase the reliability of product [8].

In the CL-PLM with the help of intelligent products, the product use data can nowadays be collected easier than before. These products which are equipped with sensors and embedded technologies, can capture and transmit data when they are under operation. Yet, the modern use data has complex characteristics. These characteristics include the followings; they come from various sources and different formats. They are being produced with a high speed and their amount is larger than the use data gathered previously from the product. Therefore, in order to improve the product and the product lifecycle processes with these data, discussing the use of data analysis in the product lifecycle is necessary.

From the other hand, mentioned characteristics of modern use data, collected by smart products are similar to “Big data” characteristics. Big data is the data which has high volume, high speed of generation (velocity of generation) and consists of variety of source and formats [20]. In this paper, we discuss the product use data from the perspective of big data analytics. The big data analytics has not been discussed in the CL-PLM broadly so far. The aim of this study is to provide an overall view from the potential mechanism of using big data analysis in the CL-PLM. For this reason, we seek to find the reference processes. The information feedback flow from the use process to the design process is focused in this paper. The importance of this analysis is that currently, there is no instruction or standard processes which guide on how to find this information feedback and how to transform use data into useful information for this aim. Finding the reference steps in this case is very beneficial. Because it is independent from tool or technique of data capturing, analyzing and information feedback gaining. Moreover, most of the current practices in the field of PLM to model the information flows address the IT and computer science perspective. Little research has been done that focus on the processes.

Therefore, in this research we investigate the modification and integration of Cross Industry Standard Process for Data Mining (CRISP-DM), which is a well-known standard process for data analysis in the CL-PLM. This is done based on the characteristics of usage data and information feedbacks between usage and design phase. This paper is organized as follows. Section 2 describes the state of the art. Section 3 explains the research approach for integrating the data analysis to the CL-PLM. Section 4 provides a brief discussion and in section 5 the conclusion is presented.

2 Related Work

2.1 CL-PLM and Information Feedbacks to Improve Design

Traditionally, the product use data can be gathered mainly by methods such as questionnaire and interview from the customers or by analyzing the failures happened to the products from maintenance or warranty reports. This information is later applied to improve the design reliability. However, this kind of feedback generation takes long until it turns to actionable information for the designers [1]. Also, the designers should apply a lot of assumptions about the product's condition of the use. However, by development of the concept of CL-PLM, accessibility to usage data is increased. In addition, processing of these data and new methods of getting information feedbacks gained significant attention. One reason is that now it is possible to achieve a complete view of the product usage instead of only using the data from usage measurements [5].

Several researchers addressed the information feedback from the usage data in the CL-PLM by data analysis techniques. For example, [1] discussed the integration of data analysis methods in the production phase for manufacturing of steel. [13] considered the feedback generation from product use when data for several instances of the same product should be summarized and extracted. [13] used the Bayesian method as the technique to generate information feedbacks from usage data for the aim of improving the product design. [1] studied integration of usage data from the condition monitoring system to PLM systems. [16] made a similar research on the data of condition monitoring systems for a conveyor belt and proposed a methodology to integrate the results in BOL. From the literature it can be observed that data analysis plays an important role in transforming the usage data to relevant information feedbacks for the product design.

In order to gain better information feedbacks, it is important to identify the source and understand the characteristics of usage data [36]. This knowledge can help to find a suitable data analysis method for generating feedbacks. More description in this regard is provided in section 2.2.

2.2 Product Usage Data and its Characteristics

The usage data in the CL-PLM can be gathered by smart products. In this paper smart products are referred to consumer products which are equipped with sensors, RFID or embedded technologies. They can collect the information about their status and use condition [19] & [31].

The sensors installed on the devices, can stream the data such the environmental condition, status of product and history of changes, type of the use and performance of product.

There are also other sources of product use data. The data which can be gathered from mobile applications, social media and websites. These type of data can show the user's opinions about the product or problems with the product.

All these data sources, have specific characteristics. They can be collected every few minutes. They are being generated very fast. For example, in the case of sensor data, the measurements can be done every few minutes. They have various formats and characteristics. For example, the sensor can be presented in log files or excel sheets, while the text data from maintenance reports is unstructured and cannot be presented well with the excel sheets. Moreover, usually they contain not only one measurement, but also a batch of data for every measurement interval. Therefore, when we consider the amount of the data and the speed of their generation we are exposed to a big amount of data. It can be said that the product use data has the characteristics of variety, velocity and volume (3V), similar to the characteristics in the context of big data analytics. More explanation is provided in section 2.3.

2.3 Big Data Analytics to Support Getting Information Feedbacks from Usage Data

Big data is considered as *“high-volume, high-velocity and high-variety data that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”* [20]. The data analytics is part of big data technology which aims to convert the data into useful information. These information has the potential to provide insight for the decision making. For example, in the maintenance it can help to find the failures before they happen. In spite of slight differences, the terms big data analytics, data mining and data analysis are used interchangeably in this paper.

Data analysis has been seen as a complementary service and not as a main module in the CL-PLM. However, regarding the importance of usage data and its potential to improve lifecycle activities, it is worthy that usage data be discussed and more investigated. One aspect which needs attention is to find a uniform and standard guideline for the use of big data analysis in the CL-PLM. Applying a standard can have several advantages for CL-PLM. For example, it can act as a guideline. So, it reduces the need for the high skilled people for analysis of data. Also it leads to time and cost saving. From the theoretical perspective, it offers stable model development for the problems in the lifecycle because it is a generic model and can be used apart from the tools and techniques used for modelling.

In the next section CRISP-DM as a standard for processes of data analytics and data science is explained. Additionally, its applicability to the CL-PLM is tested.

2.4 CRISP-DM

“Cross Industry Standard Process - for Data Mining” (CRISP-DM) is a data analysis process standard that describes commonly used approaches for performing analysis of data when the volume of the data is high. It is applicable in various industries. This standard was founded by SPSS in a cooperative project, where Daimler Chrysler, was

also a shareholder [23]. Currently, this process model is supported by IBM. It is one of the most widely used standard by the data mining practitioners. This model consists of different levels of abstraction. Fig. 1 shows the data mining methodology based on CRISP-DM processes and sequences. In this figure the high level processes are shown. The business understanding shows the requirements of the data analysis problem from business perspective. It emphasizes on understanding the goal of data analysis and recognizing the aspects of problem very well. The data understanding process includes the initial data collection and becoming familiar with the data, its variables and dimensions. The data preparation contains the filtering, aggregating, selecting the parameters and constructing a subset of data suitable for the analysis. The modelling phase different models are fitted to the data and the optimal values are found. The relevant models covered are the machine learning, data mining and statistical analysis. At the evaluation stage the goodness of the model and the outcome gained from it is assessed. At last, during the deployment phase the knowledge gained from the modeling is discussed with the user and applied to the problem in action [37].

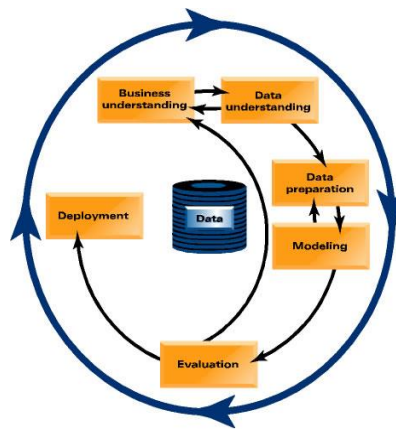


Fig.1. CRISP-DM processes and cycle [23]

The aim of selection of CRISP-DM is, that it can cover analysis of data when the data has high volumes. Therefore, it can be suitable for modelling the use data. However, it does not take into account all the characteristics of product use data such as variety of sources and fast generation. For example, the data can come from the sensor also from the web. There are no instructions in the CRISP-DM for handling variety of formats specially the unstructured data. Moreover, it does not take the data veracity into account. Similar problems exist when it comes to data that are generated very fast for example by sensors. These aspects have not been considered in this standard. In the next section in order to leverage the limitations of the standard, we investigate current available big data analytics frameworks and suggest solutions.

3 Approach: Integrating the Data Analysis in the Closed-Loop Product Lifecycle

For covering the dimensions of use data which are not included in CRISP-DM, such as variety and veracity, we go through the current available big data analytics frameworks. Twenty-three papers with big data process frameworks from the state of the art of big data analytics were selected. The papers are either published by renowned research databases or are the technical report of companies, who are active in the field of big data. For example, [30] and [14]. We determined the steps common between the frameworks. The steps regarding to conducting the analysis on data are investigated. Moreover, the steps of data analytics have been compared with the steps of CRISP-DM standard. Table 1 shows this comparison.

Table 1. Assessment of big data analytics frameworks and comparison with CRISP-DM steps

Framework name	Business understanding	Data understanding	Preparation	Modelling	Evaluation	Deployment	Other processes
(Big data consulting, 2012) [4]	1	1	0	1	1	0	Impact analysis, Research best practices of big data
(Business Biro, 2014)[7]	1	1	0	1	0	1	Define big data architecture
(Loshin, 2012)[27]	1	1	0	0	0	0	
(Krishnan, 2013)[25]	0	1	1	1	0	0	Data acquisition , data storage, big data visualization
(Philip Chen & Zhang, 2014)[30]	0	0	1	1	0	0	Data storage, decision making, Big data visualization
(Reffat, et al., 2006)[32]	0	0	1	1	0	0	Data storage
(Dayal, et al., 2009) [11] a	0	1	1	1	0	0	Define big data architecture, data generation, Data warehousing (storage)
(Dayal, et al., 2009) b	1	0	0	1	1	1	Data flow integration
(Dayal, et al., 2009) c	0	0	1	1	1	0	Unstructured data modeling
(Assunção, et al., 2015) [3]	1	1	1	1	1	1	Data source, big data visualization
(Pääkkönen & Pakkala, 2015) [29]	0	1	1	1	1	1	Data generation, data management, data extraction (acquisition), storage
(Vera-baquero, et al., 2014) [35]	0	0	1	1	1	1	Control and monitor the results
(Lavalle, et al., 2010) [26]	1	1	0	1	0	1	
(Ren & Zhao, 2015) [33]	0	1	1	1	0	1	Data acquisition, data storage, data management

(Gandomi & Haider, 2015) [20]	0	0	1	1	0	1	Data acquisition and record (storage), data management
(Zhan, et al., 2013) [38]	0	0	1	1	0	0	Data generation
(Catley, et al., 2009)[9]	1	1	1	1	0	1	
(Folorunso & Ogunde, 2004)[18]	0	0	0	1	1	1	Decision maker, BPR, new uses in process
(Fernández, et al., 2014) [17]	0	0	1	1	0	1	Data management, parallel computing, data storage
(Accenture, 2013)[2]	1	0	1	1	1	1	Matrices, insight validation, Business review cycle
(Booz Allen Hamilton Inc., 2013)[6]	0	1	1	1	0	1	Data acquisition
(Diamantoulakis, et al., 2015)[12]	0	0	1	1	0	0	data source (generation), distributed processing
(Dutta & Bose, 2015)[14]	1	0	1	1	1	1	Data acquisition, data visualization, integration with IT systems
(Hackathorn, 2013)[21]	0	1	1	1	1	1	Automate and operationalized, data acquisition
(Hashem, et al., 2014) [22]	0	0	1	1	1	0	Data source, data visualization
Sum	9	12	19	24	11	15	

In table 1, the frameworks are listed. They are compared with the CRISP-DM (column 2 to 7). Column 2 to 7 shows the phases of CRISP-DM, including business understanding, data understanding and etc. If the framework supports any of these phases, it is marked with 1. Otherwise, if the process is not considered in the framework it is marked with 0.

As mentioned before, the frameworks which cover the aspects of data variety and veracity were selected. Therefore, complementary processes exist in the framework, which is not included in the CRISP-DM phases. The complementary processes are presented in the last column of the table under the name “other processes”. At the end the frequency of observed processes is calculated.

Table 2. Frequency of observed processes. Results from 25 big data frameworks

Big data analysis steps	Frequency observed in the frameworks
Data generation	6
Data acquisition	6
Data storage	6
Business understanding	9
Data understanding	12
Data management	3

Parallel computing	2
Preparation	19
Modelling	20
Evaluation	11
Visualization	5
Deployment	15

Table 2 shows the steps of processing big data, from initial data generation to gaining the useful knowledge from the data and make it actionable. Also the frequency of their observation in the big data frameworks are reported. The steps are also relevant for product use data. These steps can be followed as a reference guide line. These steps are useful to pursue, particularly in the case a data analysis project conducted in the CL-PLM. In the following we explain more about each process.

The first process is data generation. It means first the data are produced by the smart devices. It can be in form of the measurements by the sensors, such as temperature, vibration or other parameters relevant to the functionality of the product or its condition of use. Afterwards, the data are acquired from the smart product and put in to the storage area or to the analytical system. The importance of data acquisition is, to find all the information about the product usage, some sources of data need to be automatically extracted from the internet. The new data acquisition procedures and tools has been developed in the recent years to fulfill this need.

Data storage process come afterwards. The data storage has not been considered as an independent and major process in CRISP-DM. A reason can be in the CRISP-DM only storage in form of relational databases has been considered. In other words, storage of structured data. However, the data which are relevant for product use are partly in form structured data. They are also in other formats. Such as unstructured (text) format. Therefore, the data storage in cases were data with 3V characteristics exist should be recognized as a main task. Therefore, it has a major role for CL-PLM regarding management of product use data.

One of the other processes which was not clearly discussed by CRISP-DM, but is very important relating to use data is data visualization. Visualization is one of the most effective ways to communicate the results of analysis (data behavior) with the users. Not only modelling the data matters but also how to represent is to the decision maker who wants to use it for getting the insight and making decisions based on these data. Yet, to make the data analysis suitable for CL-PLM, only the data analysis steps are not sufficient. We need to identify the steps of using the knowledge and transform it to information feedbacks to the design. For achieving this goal, in the second part of the study, a similar approach is done to identify the processes after finding initial insight from data analysis models until transforming and using the information as information feedback to the product design. In this part, analysis of 18 other frameworks was done. They were selected from the literature of improving the product design by taking into account the field data and product use, for example [15], [34]. The processes are grouped to four steps (table 3). Then the frequency of each process in the data analysis models has been calculated. The results of these analyses are presented in table 3. The

total frequency of observed process in the frameworks is reported in the second column of table 3.

Table 3: Results from 18 product design feedback frameworks

Step of product design improvement	Frequency observed in the frameworks
Root cause analysis/ tests/ experience	6
Severity analysis	2
Identification of degradation mode	3
Decision support integration	5

As illustrated in table 3, root-cause analysis process contains the methods for problem solving and transforming the insight gained from the usage data to the useful information for the designers. In this process, techniques such as Failure Mode and Effect Analysis (FMEA), Failure Tree Analysis (FTA), Failure Mode Effect and Criticality Analysis (FMECA), tests and experiments are included. In fact, these techniques are applied to the data after modeling by data analytical methods. Ranking and identifying the severity of the problem for design modification is an important step. However, in the literature cited in this survey was only observed two times as a main process. The degradation mode identification and modelling its function has also proposed by some authors. In the last process, some authors integrated the knowledge to the decision support system for giving feedback to the designers.

Discussion

This paper tried to make the first step of the standardization in CL-PLM for analysis of the usage data collected from the smart products and introduce how a solution in this respect could look like. In this respect, we reflect the applicability of CRISP-DM. Moreover, we investigated the standard processes of usage data analysis and information feedback to product design. Some issues which need attention are listed as follows.

The Scale of Data Analysis: In the CL-PLM we can model the information flows regarding one product item or class. The standard processes of the data analysis and feedback generation (table 2 &3), can be the same for all these categories. However, based on the scale, the focus of main processes can vary. For example, in the case of item-level data analysis enabling the track and trace of the product matters. In the case of complex engineering products, it can be important to model the interaction between several constituted modules (parts) of the product with each other. As an instance, in the case of degradation analysis, the effect that one faulty part can make for the other parts near it. In the case of mass produced products, the advances in the field of IoT,

summarizing the knowledge gained from analyzing the behavior of several products and connecting the products should be considered.

The Uncertainties of Mapping. Methodologies for analyzing the data from intelligent products is still under development. The best practices for analyzing the product use data in the CL-PLM are not still specified. Consequently, standardization of process for data analysis on use data in the CL-PLM is on its early phase. However, in this research first step is made in this regards. This was done through analyzing the relevant literature and suggesting a current applicable standard.

Open Issues. We need to have approaches to deal with increase of product usage data. Also there are considerations for the modeling in action. For example, the availability of complete data. Specifically, when the use condition is captured but not all the necessary data which affect the problem under study has been captured. Aspects of data analysis, for example the difficulties in data storage when the usage data volume is very high and need distributed storage or the use of cloud services, also analysis of unstructured data still need research.

Conclusion

In this research analyzing the product use data for improving the design activities was addressed from the view point of big data analytics. First, the characteristics of new sources of product use data was described. Afterwards, a relevant methodology called CRISP-DM, from the field of data science, was introduced to the CL-PLM and its integration discussed. The steps of data analysis to information feedback generation for the design was the outcome of this paper. These steps can be a guide for the ones who want to apply data analysis in CL-PLM. The future work includes further investigation of the applicability of CRISP-DM for other types of information feedbacks in the lifecycle. Such as, feedback from the use data to the production or to the end of life phase. In addition, the proposed standard processes should be tested with case studies.

Acknowledgment. This research is supported by the DAAD-GSSP (Deutsche Akademischer Austausch Dienst) scholarship.

References

1. Abramovici, M., Fathi, M., Holland, A. & Neubach, M.: Integration of product use information into PLM. Proceedings of the 15th CIRP International Conference on Life Cycle Engineering, The University of New South Wales, Sydney, Australia (2008)
2. Accenture: Analytics in Action: Breakthroughs and Barriers on the Journey to ROI, Technical Report, Accenture (2013)
3. Assunção, M. D. et al.: Big Data Computing and Clouds: Trends and Future Directions. Journal of Parallel Distributed Computing, vol. 79, 3–15 (2015)
4. Big data consulting, <https://bigdataconsulting.wordpress.com/consulting/> (2012)

5. Bluenose, <http://www.bluenose.com/blog/6-ways-to-increase-revenue-with-product-usage-data> (2016)
6. Booz Allen Hamilton Inc.: The Field Guide to Data Science, 2nd Ed. (2013)
7. Business Biro, <http://www.businessbriio.com/big-data.html> (2014)
8. Carlson, J. & Murphy, R: Reliability analysis of mobile robots. In: Proceedings of IEEE Robotics and Automation (ICRA'03), pp.274-281 (2003)
9. Catley, C., Smith, K., McGregor, C. & Tracy, M.: Extending CRISP-DM to Incorporate Temporal Data Mining of Multi- Dimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study. In: IEEE International Symposium on Computer-Based Medical Systems, pp 1-5, (2009)
10. CL2M, <http://promise-innovation.com/cl2m> (2008)
11. Dayal, U., Castellanos, M., Simitsis, A. & Wilkinson, K.: Data Integration Flows for Business Intelligence. ACM, Saint Petersburg, Russia, pp. 1-11 (2009)
12. Diamantoulakis, P., Kapinas, V. & Karagiannidis, G.: Big Data Analytics for Dynamic Energy Management in Smart Grids. In: Big Data, Analytics, and High-Performance Computing, vol. 2(3), pp. 94–101 (2015)
13. Dienst, S., Ansari-Ch, F., Holland, A. & Fathi, M.: Necessity of using Dynamic Bayesian Networks for feedback analysis into product development. In: IEEE International Conference on Systems Man and Cybernetics, pp. 939-946 (2010)
14. Dutta, D. & Bose, I.: Managing a Big Data project: The case of Ramco Cements Limited. In: International Journal of Production Economics, vol. 165, pp. 293-306 (2015)
15. Fathi, M. & Holland, A.: Knowledge-based feedback integration to facilitate sustainable product innovation. In: IEEE conference on Emerging Technologies & Factory Automation, ETFA 2009. IEEE, 2009 Mallorca, pp. 1-8 (2009)
16. Fathi, M., Holland, A., Abramovici, M. & Neubach, M.: Advanced Condition Monitoring Services in Product Lifecycle Management. Las Vegas, IL, In: IEEE, pp. 245-250 (2007)
17. Fernández, A. et al.: Pattern Recognition in Latin America in the “Big Data” Era. In: Pattern Recognition, vol. 48(4), pp. 1185–1196 (2014)
18. Folorunso, O. & Ogunde, A.: Data Mining as a Technique for Knowledge Management in Business Process Redesign. In: Electronic Journal of Knowledge Management, vol. 1(2), pp. 33-44 (2004)
19. Främling, K. & Nyman, J.: Information architecture for Intelligent Products in the Internet of Things. In: Beyond Business Logistics proceedings of NOFOMA, pp. 224-229 (2008)
20. Gandomi, A. & Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. In: International Journal of Information Management, vol. 35, pp. 137-144 (2015)
21. Hackathorn, R.: Evolution of Big Data Analytics: Experiences with Teradata Aster and Apache Hadoop, Bolder Technology Inc. (2013)
22. Hashem, I. A. T. et al.: The rise of “Big Data” on cloud computing: Review and open research issues. In: Information Systems, vol. 47, pp. 98-115 (2014)
23. IBM, Managing big data for smart grids and smart meters, Somers, IBM Corporation, New York (2012)
24. Kiritsis, D.: Closed-loop PLM for intelligent products in the era of the Internet of things. In: CAD Computer Aided Design, vol. 43(5), pp. 479-501 (2011)
25. Krishnan, K.: Big Data Processing Architectures. In: Data Warehousing in the Age of Big Data. Waltham, Morgan Kaufmann, MA, USA, pp. 29-43 (2013)

-
26. Lavalley, S. et al.: Analytics: The New Path to Value. In: MIT Sloan Management Review, IBM Corporation (2010)
 27. Liu, W., Zeng, Y., Maletz, M. & Brisson, D.: Product Lifecycle Management: A Review. In: International Design Engineering Technical Conferences and Computers and Information in Engineering, San Diego, USA, pp. 1213-1225 (2009)
 28. Loshin, D.: Business Processes and Information Flow. In: Business intelligence.: the savvy manager's guide. Newnes, pp. 77-90 (2012)
 29. Pääkkönen, P. & Pakkala, D.: Reference Architecture and Classification of Technologies. In: Products and Services for Big Data Systems. Big Data Research, vol. 2(4), pp. 166-186 (2015)
 30. Philip Chen, C. & Zhang, C.-Y.: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. In: Information Sciences, vol. 275, pp. 314–347 (2014)
 31. Ranasinghe, D., Harrison, M., Främling, K. & McFarlane, D.: Enabling through life product-instance management: Solutions and challenges. In: Journal of Network and Computer Applications, vol. 34, pp. 1015-1031 (2011)
 32. Reffat, R., Gero, J. & Peng, W.: Using data mining on building maintenance during the building life cycle. In: Proceedings of the 38th Australian & New Zealand Architectural Science Association (ANZASCA) Conference, University of Tasmania, pp. 91-97 (2006)
 33. Ren, S. & Zhao, X.: *A predictive maintenance method for products based on big data analysis*. In: International Conference on Materials Engineering and Information Technology Applications (MEITA 2015), pp. 385-390 (2015)
 34. Shin, J. H., Kiritsis, D. & Xirouchakis, P.: Design modification supporting method based on product usage data in closed-loop PLM. In: International Journal of Computer Integrated Manufacturing, vol. 28(6), pp. 551-568 (2015)
 35. Vera-baquero, A., Colomo-palacios, R. & Molloy, O.: Towards a process to guide Big Data based Decision Support Systems for Business Processes. Procedia Technology, vol. 16, pp. 11-21 (2014)
 36. Wellsandt, S., Hribernik, K. & Thoben, K.-D.: Sources and characteristics of information about product use. In: Procedia CIRP 36, pp. 242-247 (2015)
 37. Yun, Z., Weihua, L. & Yang, C.: Applying Balanced ScoreCard Strategic Performance Management to CRISP-DM. In: Information Science, Electronics and Electrical Engineering (ISEEE), pp. 2009 – 2014 (2014)
 38. Zhan, J. et al.: BDGS: A Scalable Big Data Generator Suite in Big Data Benchmarking. In: *Advancing Big Data Benchmarks*, pp. 138-154 (2013)