Associative Data-Process Model in Manufacturing Systems: Application Case in Automotive Industry

Letícia Alves dos Santos ROSOLEM, Alexandre da Costa CASTRO, Fernando DESCHAMPS and Edson Pinheiro de LIMA
Pontifical Catholic University of Paraná, Curitiba, Brazil

Abstract. Considering the advent of Industry 4.0 and development of Cyber-Physical Systems, a large amount of data has been collected in production systems. Available business process recommendations, presented in Business Process Management Notation and Case Management Model Notation frameworks, have been enough to represent the systemic and relational characteristics of production systems and their components. Recommendations for data life-cycle management do not consider explicitly combining the elements of the modeled process and the representative data of the production system assets. In fact, process flows can represent data generated or consumed but do not have an explicit associated action to their data: generation, transformation and consumption. From a Performance Analytics perspective, approaching data and processes in an associated way can be an advantageous practice in management and decision-making. This study uses one investigative experience in the automotive industry, from two aspects: the As-Is mapping of a segment of the measurement system after the welding step, and the design and implementation of Big Data Analytics architecture related to the same process. The result is a proposal of an associative framework between processes and related data, which are following the recommendations of currently applied frameworks for Business Process Management and Big Data Analytics.

Keywords. Big Data Analytics, Business Process Management, Manufacturing Process, BDA Architecture, Transdisciplinary

Introduction

With the advent of digital transformation and new technologies, the data generated in processes has increased in their production systems. To ensure greater competitiveness, quality and efficiency, companies have adapted to these new data analysis processes through CPS and IoT, capable of consuming and generating data. Therefore, there was a need to better understand and apply the data analysis processes and the relationship with the respective stages of the production process. There are studies related to the development of processes for data analysis based on industrial processes such as maintenance presented by [1][2], and research and practice related to Big Data architecture [3][4][5]. However, little aspects has been studied about the correlation between each domain, or the its contributions which characterizes the transdisciplinarity phenomenon.

The Business Process Management (BPM) area has a solid framework stabilized by a good theoretical framework and stabilization through continuous practice. The components of the framework adequately express the operational aspects (e.g.,

sequences, transition conditions, etc.) and attributes inherent to their participation in the process (e.g., identification, names, other attributes). Another aspect to be highlighted is the adoption of the standard by practitioners in industry and research, concerning to representing and improving processes. In one of its derivations, the CMMN stands out allowing a more flexible representation of the process, in an appropriate way and without losing its stability as a framework notation. The root domains related to Big Data (e.g., engineering, processes, applied mathematics) are characterized by the strength of their definitions proven by years of practice and research. A lot has evolved in the last 20 years, allowing the stability of the resources that support the current Big Data infrastructure, mainly those related to Cloud Services. Research and practice suggests an approximation between components related to the Big Data architecture and the application object of process management.

The present study has as main objective to identify if the process management can contribute in an assertive way with the structuring of a Big Data architecture, without a mischaracterization of the needs of the latter one. To this end, a case study was developed in the automotive industry related to the Big Data infrastructure of a Car Frame production system in which weld measuring data is obtained, stored, transformed and used in the production line management decision process.

1. Literature Review – Business Process Management and Data analytics

1.1. Business Process Management

To develop a business process it is important that the model describes well the steps that make up the development of a process. Therefore, this model must understand all its activities, users and attributes to achieve milestones. Process models for data analysis need to have greater flexibility in their flow and automation of their processes. Hence, to develop a process considering the data flow, allowing its flexibility, we consider the PAIS methodology (Process Aware Information Systems). According to [6], PAIS aims to meet the demands of information systems through a logical process that manages, executes and analyzes processes of an organization based on specific models seeking agility and flexibility. When performing a PAIS based on models of executable processes, it is important to consider the process variables, they may have different characteristics and needs. For example, we must consider that there are well-structured processes whose behavior can be totally pre-specified. As well as, highly dynamic processes [7].

According to [7], it is important to develop the business model considering a high level of abstraction so that they serve as a basis for analysis, simulation and processes visualization. Therefore, the model must understand all the activities of the process and its attributes, as well as the control and flow of data between the activities so that we can develop a more detailed process. For the development of the process we will use the BPMN notation (Business Process Management Notation), through the method known as Case Management Model and Notation (CMMN). According to [8] [9] the CMMN is an unstructured data centered management process language with a high degree of flexibility, which means that the main and permanent assets are the data. In addition, the notation has a central concept that contains all of the process data information, also is a type of business process technology that uses the flow of control to describe the business process. The construction of process management using the CMMN demonstrates that the focus is not only on the process, but it is also on the actors

who are responsible for the business objectives and use the systems as a tool to achieve those objectives. Consequently, to develop a reliable business process management models using CMMN depends more on the judgment of the actors than on the process control flow [10]. For the development of the architecture of data analysis processes, we use the study [11] where the authors recommend the use of the main areas to be considered for the development of a business process for data analysis: Data Gathering, Data Preparation, Data Analytics, Data Visualization and Data Stakeholder. In the present study we will consider these steps as the main ones for the construction of a framework for the application of the process.

1.2. Big Data Analytics

The infrastructure used for the life cycle of data related to the production line and the processing of this same data must follow the guidelines that recommend zoning, according to the characteristic and stage of data maturity in its workflow, from generation to consumption [3][12][13]. As recommended in the studies of [3][4][5][14], the zones used for processing and storing intermediate data (transient, raw, trusted and refined) must meet the minimum specifications for security, retention, content, and data sensitivity. The transient zone must accommodate the data recently ingested from the operation without any transformation that would alter its originality. The goal of the agent responsible for this intake is to move the data from the source and, so on, store it in this zone. At this stage, the *Velocity* characteristic of Big Data environments is present, which justifies the mission defined for the data transition agent: with speed, obtain the data and store it immediately. The Raw zone is now accommodating basic data for further processing with a certain degree of adjustment and cleanliness. In the model proposed by [11], cleaning, harmonization and other Data Wrangling tasks can occur according to the needs or business rule. The agents who operate in this area, follow the same compliance guidelines and adapt to the flexibility proposed in the model. The trusted zone is for store data without its sensitive part (which does not apply in all the cases). Finally, the Refined zone is intended to contain the refined and harmonized data with meaning to the aspects of the business and operation, ready to be consumed by decision support systems and business analytics dashboards.

2. Conducting the research

The development of the study was carried out through technical visits and experience in the production operation plant. Diagnostics are concentrated in two major areas: processes and data.

2.1. As-Is Process

The process developed in the framing line occurs in 6 different lines where the data capture takes place in 4 points. Figure 1 represents a step in the process that consists of welding the plates that make up the product. The first stage is the reception of the car Frame, where the internal parts of the product are added, these parts are glued and their tabs are folded, these tasks are performed by machines on the line. Then, the application of the anti-noise sealant is carried out manually. After that, the frame geometry is

executed as well as some spot welds in the same stage. After calibration, two more spot welds are carried out using mechanical arms. Finally, there is the verification of the weld geometry using image sensors that calculate the size and geometric location of the weld point. This step also generates a validation of the weld point, to check if it is in accordance with the reference parameters, allowing to obtain the data for analysis.

After completing the steps described above, it is possible to make decisions based on the data analysis: (i) If the welding point is in conformity, the Frame is sent to the next step; (ii) If the welding point is not in conformity, a user must check the error alert, define a correction for this error, and share the decision with a multidisciplinary team. This last decision in particular can result in the complete compromise of the Frame due to the flaws identified at the weld points. Otherwise, the correction will be carried out and the respective unit will be sent to the next step.
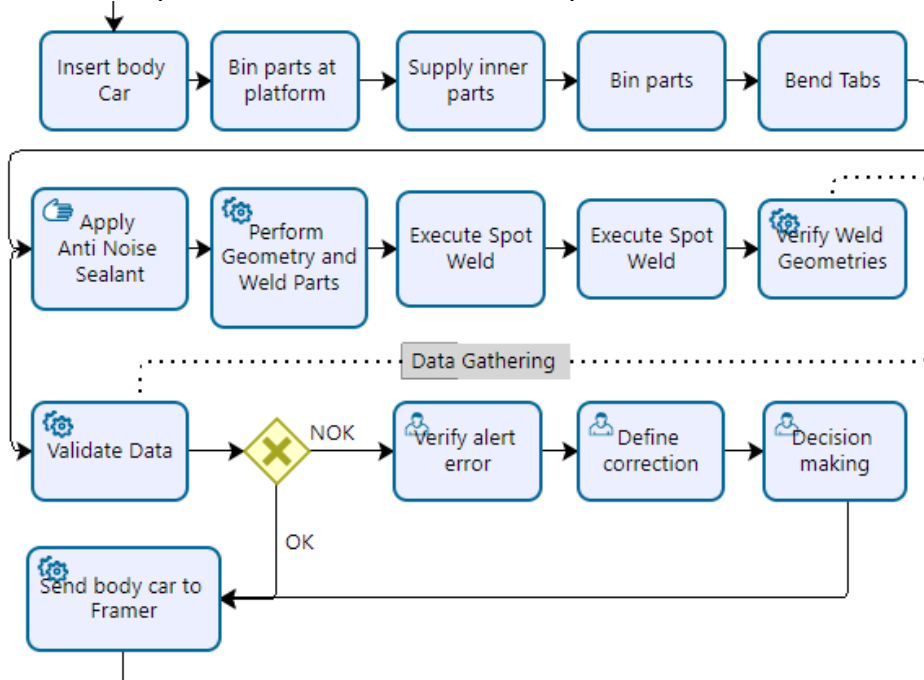

Figure 1. Framing Line Process (BPMN)

2.2. Data Life-cycle

The data obtained in the welding measurement process are stored in an intermediate area in tabular files, containing information about the unit and the corresponding coordinates measured. As for the coordinate information, it was possible to observe that not only the deviations are reported (e.g. relative deviations) but also the absolute deviation information in relation to the measurement reference points. This characteristic makes the files containing the measurement information, formed by several columns with floating point numerical data (float point representation). Such files are significantly larger than expected for measurements of this nature. Figure 2 shows the meta-process of the data life cycle.

These files are transferred to another area where they are subjected to the inherent Big Data architecture process envisaged. The Big Data architecture is structured in four zones: (i) transient, containing the pure data received from the measurement readings; (ii) raw, containing data already divided by measurement unit; (iii) refined, containing refined data, already harmonized and available for decision support dashboards. This structure follows the recommendations for Big Data architecture for transition zones [2] [3][4][13][14]. The movement between the zones follows transition and transformation criteria that are part of the specific business and are not subject to appreciation in this study. The trusted zone is not included in the structure because there is no sensitive data, both for the business and for the organization, that deserve its concealment or encryption.
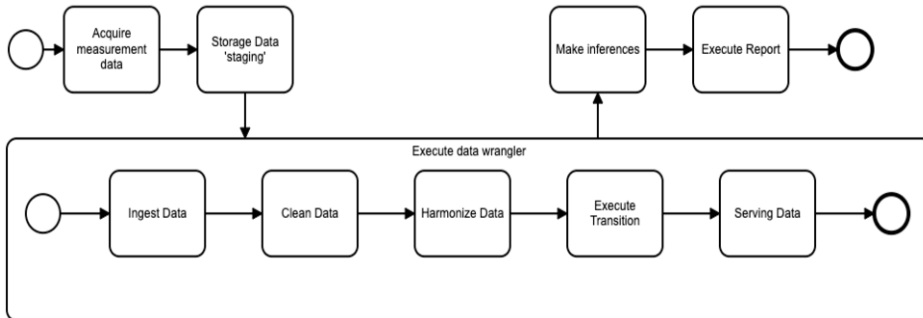

Figure 2. Meta-process of the data life cycle.

3. Diagnostic

During the research stage at the operations plant, it was possible to observe aspects related to decision making and the use of data obtained from the production line, according to the following items:

3.1. Formal and Execution Processes Correlation

The researchers identified a high degree of empiricism in the process execution and in the use of the data. Likewise, the use of data for the decision-making process follows subjectively, with a low degree of formalism in conducting transformations and transposition in the Big Data stages. Another aspect that aggravates the scenario is the low level of workflow documentation for new agents. Such a scenario is the result of the following phenomena:

- Low level of workflow formalism;
- Subjectivity of the workflow and data flow correspondence;
- Empiricism and distribution of responsibility in decision making.

3.2. Process and Data Correspondence

It was possible to identify a correspondence between the process workflow and the data transformation and transition flow between its generation and use in decision making. At

each stage of the process, there is a correspondence in the transition or transformation of data with characteristics of the business, such as, for example, the referencing of equipment or factories, part numbers and product number (PIN).

4. Main Findings

During the research process, it was possible to identify that the support structure for the production process did not have a formal definition based on the concepts of business process management, despite the existence of some schematic documentation on the process. With the design of the *As-Is* process, both researchers and practitioners had the opportunity to identify improvement points, especially regarding decision-making, as the *data journey* did not meet strictly the recommendation of the agents roles in an Big Data architecture (e.g.: cleaners, transformers, harmonizers, etc.). The CMMN process developed in this paper use the concepts of Big Data Analytics provided in source [11]. The authors describe, through the literature review, how to conduct a data analysis process and its main steps. As shown in Figure 3, the main steps are: Data Collection, Data Preparation, Data Analysis, Data Visualization and Data Stakeholder. To apply the case study, it was necessary to make some adaptations and improvements in the process presented in source [11].

      The data transformation agents depend on the performance of the actors involved in the production process, who have assumed the responsibility of structuring the data in relevant information for decision making. This characteristic shows the empiricism level adopted in the studied process and use old data in analysis. The Figure 4 presents the map of the recommended process according to CMMN notation, aligned to the recommendations produced from the observation of the current scenario.

      The process developed in Camunda Modeler Software aims to represent the relationship of the actors and their tasks related to obtaining, processing, analyzing and visualizing the data. The development of the stages was based on the studies presented by source [11] where the authors present the stages to be considered in the development of the architecture of a data analysis process. With that, it was possible to define the following structuring view of the process as a recommendation for use in the observed operation. The process begins with the project start of the digital transformation team and stakeholders. The Digital transformation team is made up by data scientist and data analysts. Also, the stakeholder team is made up by organization participants who works directly with the main process. Meetings will occur in a time based stages, which happens frequently and depends on each team.

      The first stage is Data Gathering which has the data extraction criterion as an exit criterion, however in the case study the captured data has some important characteristics to be highlighted in the its process flow, some process tasks need to be completed before closing. They are: (i) define data source, in this case the data being extracted is offline; (ii) check the data volume; (iii) check the speed of data capture; (iv) to vary the value of the data and its variety; (v) explore the data; (vi) define a tool for data storage (AWS). The second step concerning to data processing/preparation, has the exit criteria to generate the transformed and refined data and a meeting of the digital transformation team to present the first findings. At this stage, the data life cycle is an important fundamental, as it adapts the data to processing, eliminating inconsistencies, adjusting values and referring to aspects of the business (harmonization), the tool used was Alteryx Software. The third step is the data analysis where the users must define how to apply

the statistical model through the study of forecast and timeseries model, in this step the usage tool is defined to carry out the analysis (Python notebooks). Finally, the data visualization step, which consists of the definition of a data analysis tool by the user (Tableau and Power Bi), that user will check the experience and usability of the generated data, as well as view the data. At that moment, it will be possible to reach the milestone of the process, which is the BDA report, which should happen with the condition of the decision making and the meeting between the digital transformation team and the stakeholders.
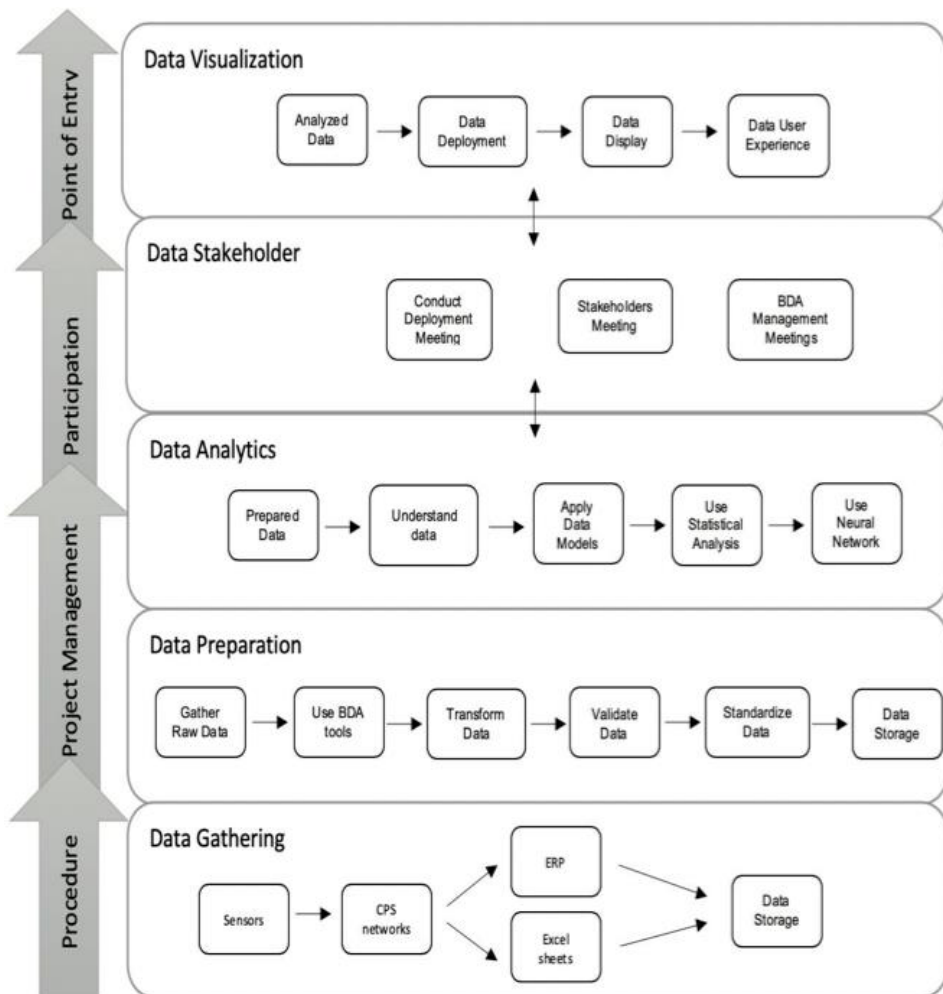
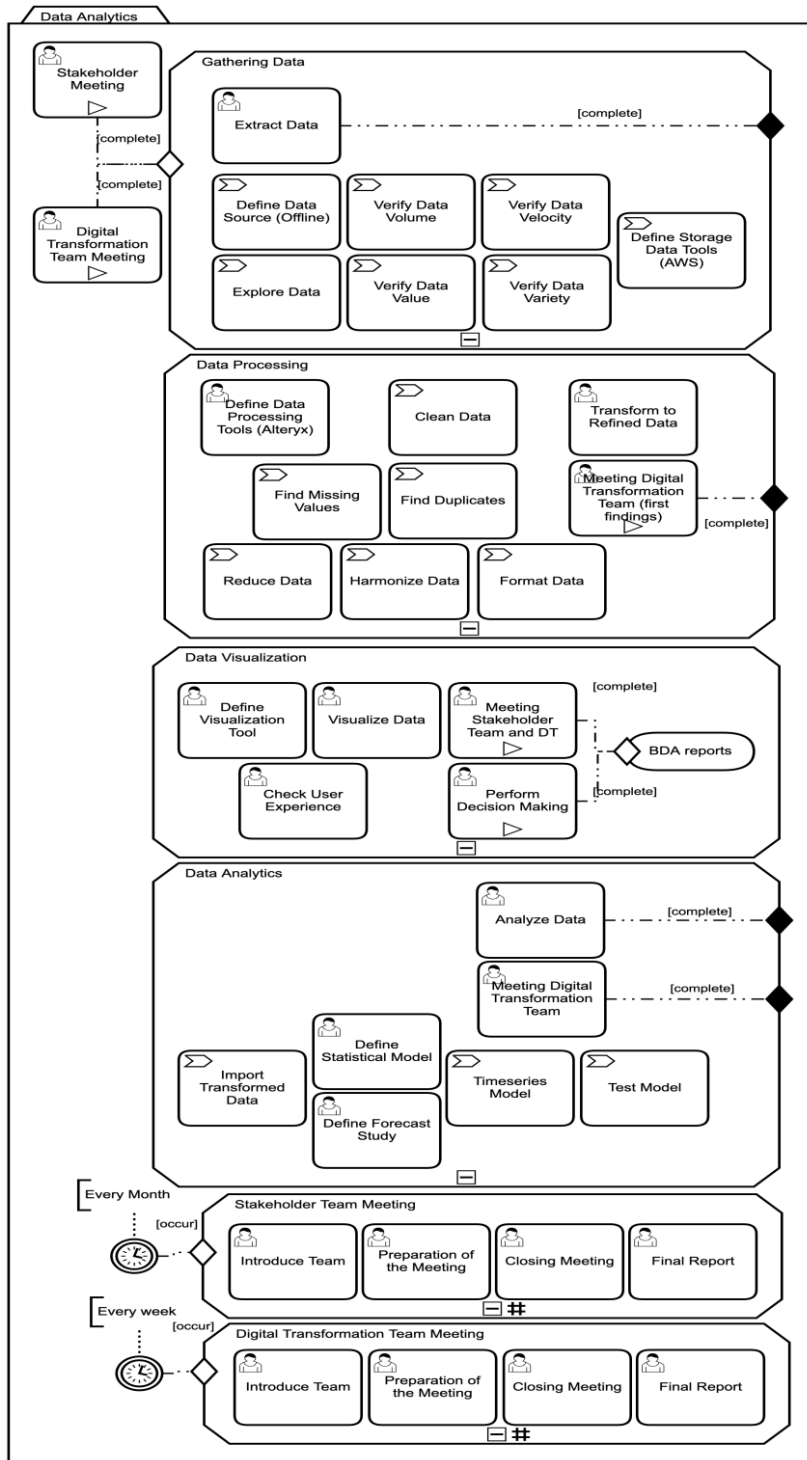**Figure 3.** Framework for Big Data Analytics Process Architecture. [11]

Figure 4. Recommended Data Analytics Process for Automotive Industry.

5. Conclusions

The data engineering process applied to big data can benefit from the recommendations and contributions of BPM, in the design, planning and definition of components, chronology and agents. Keeping the appropriate specific reserves for each domain, it was possible to perceive a correspondence between its components in addition to the way of mapping them. The actors mapped in the production process Dimensions to be addressed: actors, operational, time, information, organizational, strategic perspectives for cost and quality optimization. The experience allowed to outline a framework that flexibly guides the relationship between process and data in a prescriptive and descriptive manner, standardizing the steps of a Big Data architecture aligned to the process in which it is applied.

6. Future Works

Future works can explore in depth the relationship between the two domains, which includes derivation from CMMN notation or with BPMN. It is possible to conduct a study to the sequences of tasks related to data processing, when its mandatory. Another point that can relate the results of this study is the marking of the structured data or information components regarding their position in the process which it belongs. An exclusive decision support solution software can be designed, considering the chronology aspects and characteristics of the data obtained on the production line.

References

[1] Y. Zhang and S. Ren, Big data based analysis framework for product manufacturing and maintenance process. *IFIP Advances in Information and Communication Technology*, 2015, Vol. 460, pp. 427–435, https://doi.org/10.1007/978-3-319-22759-7_50
[2] Z. Balogh, E. Gatial, J. Barbosa, P. Leitão, and T. Matejka, Reference Architecture for a Collaborative Predictive Platform for Smart Maintenance in Manufacturing. *INES 2018 - IEEE 22nd International Conference on Intelligent Engineering Systems, Proceedings*, 2018, pp. 299–304. https://doi.org/10.1109/INES.2018.8523969.
[3] M.J. Mortenson, N.F. Doherty, S. Robinson, Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research*, Vol. 241, Issue 3, 16 March 2015, pp. 583-595.
[4] T.P. Liang, Y.H. Liu, Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, 2018, Vol. 111, pp. 2–10.
[5] M. Kiran, P. Murphy, I. Monga, J. Dugan, S.S. Baveja, Lambda architecture for cost-effective batch and speed big data processing. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2015, pp. 2785–2792.

[6] C. Ayora, V. Torres, B. Weber, M. Reichert and V. Pelechano, VIVACE: A framework for the systematic evaluation of variability support in process-aware information systems. *Information and Software Technology*, 2015, Vol. 57(1), pp. 248–276. https://doi.org/10.1016/j.infsof.2014.05.009

[7] M. Reichert and B. Weber, *Enabling flexibility in process-aware information systems: Challenges, methods, technologies*, Springer-Verlag, Berlin Heidelberg, 2012, https://doi.org/10.1007/978-3-642-30409-5

[8] M. Kocbek Bule, G. Polančič, J. Huber and G. Jošt, Semiotic clarity of Case Management Model and Notation (CMMN). *Computer Standards and Interfaces*, 2019, Vol. 66, 103354.

[9] M. Marin, R Hull and R. Vaculin, *Data Centric BPM and the Emerging Case Management Standard : A Short Survey Case Management Modeling and Notation (CMMN), An emerging OMG standard*. IBM Research - Business Process Management Workshops, 2012, 257593(257593), 24–30. Retrieved from https://acm2012.blogs.dsv.su.se/files/2012/09/2012-09-03-marin-hull-vaculin-on-emerging-OMG-case-mgmt-standard.pdf

[10] M.A. Marin, *Introduction to the Case Management Model and Notation (CMMN)*, 2016, pp. 1–23. Retrieved from http://arxiv.org/abs/1608.05011

[11] L.A.S. Rosolem, F. Deschamps, E.P. Lima and J.M.A.P. Cestari, Big data analytics process architecture for manufacturing industries: a literature review. *ICPR-Americas 2020 - SPECIAL SESSION: Machine Learning and Big Data in industrial processes*, 2020.

[12] R. Wallis et al., Intelligent utilization of digital manufacturing data in modern product emergence processes, *Advances in Transdisciplinary Engineering*, 2014, Vol. 1, pp. 261-270.

[13] J. Sheng, J. Amankwah-Amoah, X. Wang, A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 2017, Vol. 191, pp. 97–112.

[14] S. Deshmane, *Using the Lambda Architecture on a Big Data Platform to Improve Mobile Campaign Management BIG DATA*. Talentica Software, 2015, https://www.talentica.com/wp-content/uploads/2018/01/Big-Data-Using-Lambda-Architecture.pdf, Accessed July 3, 2021.