

Study of the Use of Data Mining in Modeling Non-standard Processes in a Higher Education Institution



Diogo Rocha Ferreira Maia, Renato de Campos,
and José de Souza Rodrigues

Abstract Business process management (BPM) is an administration discipline to design, improve and manage processes, and consists of several phases. One of them is the modeling of the current process (as is), during which the process is mapped as it is. One tool that can be used in the process discovery and modeling phase is data mining in information systems. Data mining techniques allow a series of applications in organizations, either as a verification process or as a discovery process. The purpose of this article is to analyze the use of data mining in the modeling stage of complex and not well structured processes when applying BPM to improve processes at a Federal Institute of Higher Education in Brazil. Due to the fact that the nature of the products and services requested in the institute's system is very varied, the processes are not standardized, and the use of the information system is not done in a disciplined way, data mining was not enough to identify the complete processes and the flow of activities in detail. Although it was not enough, data mining facilitated the definition of instance flows and made it possible to detect critical points in the process, such as points with a long duration of time. With data mining, the analysis time, meetings and modeling effort of the analysts and users were reduced.

Keywords Business process management · BPM · Data mining · Non-standard processes · Higher education

1 Introduction

Organizations that operate in the current complex and turbulent environment need to implement changes in their structures and processes. Therefore, new management approaches are necessary to allow organizations to grow and increase their competitiveness [1–3].

D. R. F. Maia (✉)

Federal Institute of Education of Roraima, Boa Vista, RR, Brazil

e-mail: diogo.maia@ifrr.edu.br

R. de Campos · J. de Souza Rodrigues

São Paulo State University, Bauru, SP, Brazil

© The Editor(s) (if applicable) and The Author(s), under exclusive license

339

to Springer Nature Switzerland AG 2020

A. M. T. Thomé et al. (eds.), *Industrial Engineering and Operations Management*,

Springer Proceedings in Mathematics & Statistics 337,

https://doi.org/10.1007/978-3-030-56920-4_27

Being part of the management task, it is necessary to describe the structure and implementation of business processes to continually improve them. Business Process Management (BPM) is an approach capable of supporting both the description and the improvement of processes. Both can be performed incrementally or radically [4].

BPM consists of several phases. One of them is the modeling of the current process (as is), during which the process is mapped as it is. Then the process is analyzed and improved, to design the future process (to be) [5].

One tool that can be used in the process discovery and modeling phase is data mining in information systems. Data mining techniques allow a series of applications in public administration or private companies, either as a verification process or as a discovery process [6].

The purpose of this article is to analyze the use of data mining in the modeling stage of complex and not well structured processes, specifically applying BPM to improve processes at a Federal Institute of Higher Education in Brazil.

2 Theoretical Reference

The way the processes are designed and carried out affects both the quality of the service and the efficiency with which the service is delivered. An organization can outperform another that offers similar services if it has better processes and performance. This holds true not only in customer-facing processes, but also in internal processes [7].

In the last decades, there has been a growing interest in BPM due to its ability to help organizations increase productivity by achieving operational excellence and reducing costs [8].

Research in this field originated in computer science, administration and information systems, and has resulted in a multitude of models, methods and tools that support the design, approval, management and analysis of business processes [8]. For Aalst, La Rosa and Santoro [9], “BPM is the discipline that combines approaches to the design, execution, control, measurement and optimization of business processes”.

BPM life cycles are steps and activities that must be followed to conduct BPM projects. Theoretical and empirical studies show differences in the number of steps and activities that must be carried out to promote BPM [10].

The BPM discipline can be seen as a continuous cycle that involves a series of phases, such as process identification; process modeling; process analysis; process redesign; process implementation; and process monitoring and control [7].

Business process modeling emerged from the need to explain and communicate business processes in an organization, making them easier for business users to understand. These users range from business analysts, who design the initial drafts of the processes, to the technical developers responsible for implementing them, and even the business team that implement and monitor such processes [11].

According to Aalst [12], creating models is a difficult and error-prone task. He cites some typical modeling errors, such as: the model describes only one version of reality; the model is unable to adequately capture human behavior; and the model is at the wrong level of abstraction.

The practice of process modeling emerged as a key instrument to allow decision making in the context of the analysis and design of information systems with process recognition [13].

Process mining aims to build a process model using an event log and a process discovery algorithm; such a technique has been applied for the discovery, modeling and improvement of processes [14].

Tiwari et al. [15] explain that among the main drivers of increased business process mining is the need for companies to learn more about how their processes operate.

The goal of process mining is to leverage event data to understand how an organization works. With process mining, it is possible to discover the sequence of tasks that are performed in a given business process as well as the interactions that occur between the participants in that process [16].

Business process mining can be used as a tool to discover how people drive processes in the real world. Dustdar, Hoffmann and Aalst [17] distinguish three different perspectives in the mining of business processes:

- Process perspective: focuses on the ordering of activities (the process control flow). The goal is to find an acceptable representation of all possible paths within the process. These paths can be expressed in terms of a process model (for example, Petri net or event-oriented process chain);
- Organizational perspective: focuses on originators within a process, that is, the people and roles that are involved and how they are related. This approach can be used to portray the roles and relationships between individuals in a process in terms of a social network;
- Case perspective: takes into account the properties of the cases, that is, attributes that can differentiate one path through one process (case) from another.

Ferreira [16] explains that the event log can be the real log of an information system or it can be a log file created from historical data recorded in a database, for example. Whatever the source, the data in an event log must have a specific structure and must contain at least the following information:

- a case id, which identifies the instance of the process;
- a task name, which identifies the activity that was performed;
- a user name, which identifies the participant who performed the task;
- a timestamp, indicating the date and time the task was completed.

Process mining starts from event data and uses process models in several ways. For example, process models can be discovered from event data records, as well as serving as reference models, or used to design bottlenecks. In order to treat the information contained in an event's logs it is necessary to develop algorithms. The process mining algorithms for the discovery of processes transform the information from the event's logs into process models [12].

3 Research Method

This research is classified as exploratory and qualitative, and was developed as a case study in a Federal Institute of Higher Education in Brazil. The process analyzed was the Acquisition of Goods and Services. It was considered a critical process because it is extensive and complex and still involves several sectors of the organization. The process uses an information system, but it is not well structured and has a low level of standardization. This complexity is mainly due to the different types of products and services that are requested by the users of the organization.

The research consists of four stages, the first being a literature review, in order to theoretically support the research. The sources of consultation used were Emerald Insight, Scopus, Web of Science, Springer, Google and books, in addition to documentary sources, from which information was obtained from the organization.

The second stage started with the collection of data through the mining of processes in the institute's information system, describing the process of the acquisition of goods and services, which are the related activities and who are the actors involved in the process. The data were extracted in order to create an event log and refer to the sequence of the flow and processing of the goods and services acquisition processes in the period from 2014 to 2017. The data served as a basis for the development of the "as is".

In the third stage, in order to complement the information obtained with data mining and provide more quality in the modeling of the process (as is), it was necessary to hold working meetings with the objective of obtaining more details of the activities and answering questions regarding the results of the mining.

Afterwards, the analysis of the results and the conclusions of the research were carried out.

4 Development and Results

For the modeling of the processes and activity flow of the Goods and Services Acquisition process, SQL language was used to extract data from the institute's information system and to create the event log. The SQL code can be seen in Fig. 1.

After executing the SQL code, the event log generated 138 instances of the process of goods and services acquisition, and more than 3,900 procedures for the acquisition of goods and services processes were found. To facilitate the reading and demonstration of the event log result in the survey, two instances of the goods and services acquisition process were selected. An example of the event log result is described in Table 1.

Each of these participants plays a role in this process by carrying out some of the activities identified. However, when performing these activities, these agents interact with each other in a way that is not fully recorded by the system. Figure 2 shows the flow between the agents in each process activity.

```

SELECT protocolo_processo.id,
       origem.nome AS setor_origem,
       protocolo_processo.assunto,
       concat(to_char(tramite.data_encaminhamento, 'DD/MM/YY'),' ',
             to_char(tramite.data_encaminhamento, 'HH24:MI:SS')) as data_encaminhamento,
       recebimento.nome as setor_destino,
       concat(to_char(tramite.data_recebimento, 'DD/MM/YY'),' ',
             to_char(tramite.data_recebimento, 'HH24:MI:SS')) as data_recebimento

FROM protocolo_processo,

INNER JOIN protocolo_tramite tramite ON protocolo_processo.id = tramite.processo_id
INNER JOIN setor origem ON setor_origem_id = origem.id
INNER JOIN setor recebimento ON orgao_interno_recebimento_id = recebimento.id

WHERE
  status=1 AND
  search NOT ILIKE '%MIGRADO%' AND
  (search ILIKE '%compras%' OR
  protocolo_processo.search ILIKE '%aquisição%' OR
  protocolo_processo.search ILIKE '%pbs%');
    
```

Fig. 1 SQL code for extracting information about the process from the system

Table 1 Example of the result of an SQL system database query (instance 6142)

Id	Source sector	Forwarding date	Sector destination	Receiving date	Total of hours
6022	Protocol Coordination	23/06/2014 11:57	Head of Cabinet	24/06/2014 14:08	26:11:22
		15/08/2014 23:05	Administration and Planning Office	20/04/2017 18:29	23491:24:15
		24/06/2014 14:12	Administration and Planning Office	25/06/2014 15:27	25:15:15
		26/06/2014 22:35	Coordination of Contracts and Agreements	30/06/2014 18:41	92:06:18
		07/08/2014 21:16	Office of Rectory	15/08/2014 13:41	184:24:23
		20/04/2017 18:29	Accounting and Finance Coordination	29/05/2017 14:13	931:43:23

The process was modeled from the mining data of the process and represents the flow of an instance (id = 6022) of the process. It is possible to understand the flow between sectors (actors). However, it was not possible to identify the activities performed by the process participants. The event logs (id = 6022) of the purchasing process (see lines in the Table 1) represent the process procedures between the sectors of the institute. However, it was found that the process instances vary and do not allow

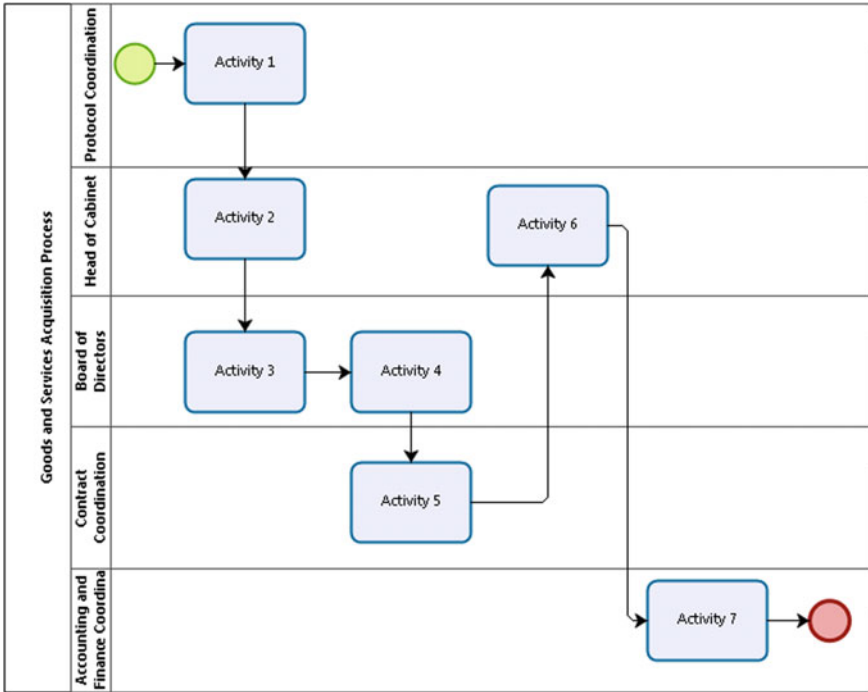


Fig. 2 Modeling process based on mining of instance 6022

a visualization of the complete interaction between all sectors. This can be evidenced by the analysis of other instances, such as that of id 6142 (Table 2).

In order to understand the flow of the process instance described in Table 2, a model was elaborated, illustrated in Fig. 3. The modeling of the instance mentioned allowed for the identification of 19 activities necessary for its completion.

Another fact resulting from process mining was the discovery of the lack of standardization in the process. This can be verified when comparing Tables 1 and 2, or Figs. 2 and 3, both resulting from the mining process. The lack of standardization is evident.

Another major discrepancy was the different interaction times of the activities sectors, recorded in the 'total hours' column in Tables 1 and 2. In a meeting with the users of the process and information system, the times of some procedures were questioned (for example, see total hours column in Tables 1 and 2). Users reported that sometimes the processing between sectors occurs only physically, leaving the information system out of date.

It was realized that this is due to the fact that the information system is vulnerable to misuse by the employees of the institute. There are no hard rules for using the system and, therefore, the real processes do not correspond to the digital processing recorded.

Table 2 Example of the result of an SQL system database query (instance 6142)

Id	Source sector	Forwarding date	Sector destination	Receiving date	Total of hours
6142	Protocol Coordination	28/08/2014 21:11	Pro Rectory of Education	02/09/2014 15:11	114:00:42
		09/09/2014 13:35	Board of Directors	15/09/2014 15:05	145:29:45
		05/09/2014 13:04	Price Research Coordination	05/09/2014 15:18	2:13:51
		22/09/2014 13:47	Protocol Coordination	22/09/2014 14:12	0:25:02
		01/10/2014 21:24	Purchasing Coordination	01/10/2014 21:41	0:16:30
		24/07/2014 13:41	Pro Rector of Administration	24/07/2014 21:03	7:21:11
		08/09/2015 14:57	Accounting and Finance Department	15/12/2015 15:20	2352:22:45
		31/10/2014 19:08	Accounting and Finance Department	04/11/2014 19:29	96:21:09
		30/10/2014 21:02	Budget Department	31/10/2014 13:06	16:04:12
		29/10/2014 12:54	Office of Rectory	30/10/2014 18:48	29:54:18
		21/08/2015 13:24	Pro Rector of Administration	25/08/2015 13:20	95:55:34
		22/09/2014 14:12	Office of Rectory	01/10/2014 18:25	220:12:27
		01/10/2014 18:25	Board of Directors	01/10/2014 21:24	2:58:45
		19/09/2014 19:12	Office of Rectory	22/09/2014 13:45	66:33:21
		24/07/2014 21:03	Price Research Coordination	07/08/2014 15:14	330:10:45
		23/07/2014 15:29	Office—Rectory	24/07/2014 13:32	22:03:23
		05/09/2014 15:18	Budget Department	09/09/2014 13:34	94:15:57
		15/09/2014 15:28	Purchasing Coordination	15/09/2014 19:14	3:45:07

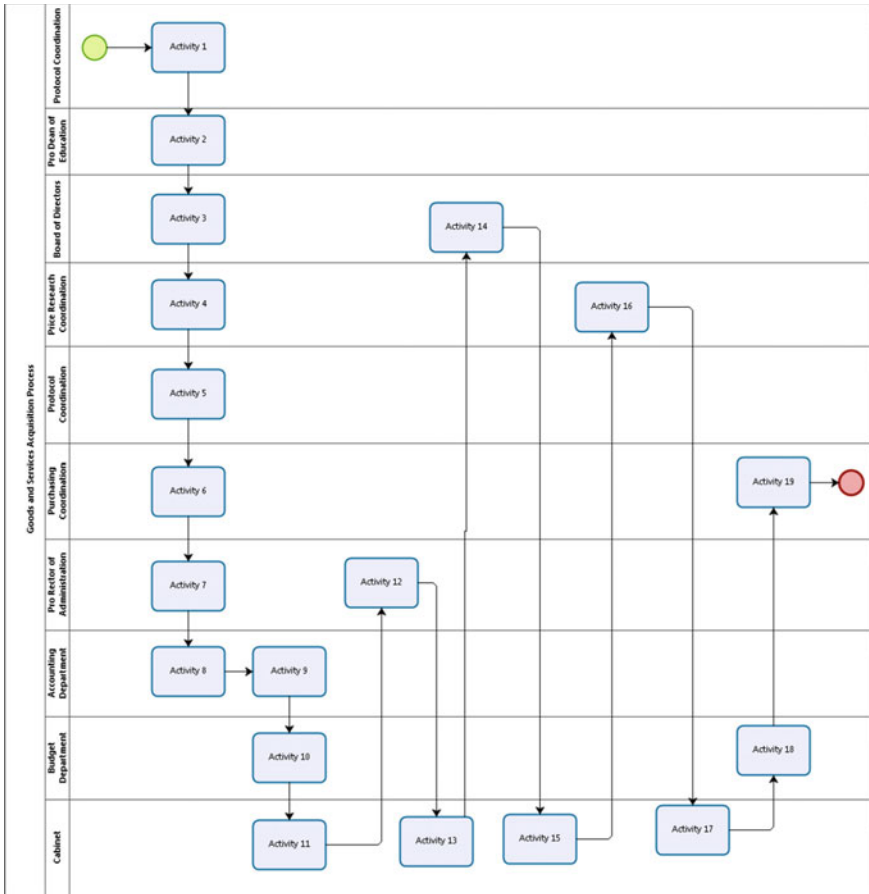


Fig. 3 Modeling process based on the mining of instance 6142

As an example, the seventh row of Table 2 describes the procedure in the Accounting and Finance Department, where the total processing time was 2,352 h and 22 min. It was found that often the processes arrive and are stopped in the sector until they receive the certificate of receipt of the equipment or until the service contracted is fully paid off with the supplier. Some services, such as the service contract for internet supply, take up to two years to finalize the total payment, as in the case of a contract with monthly payments; and after one year it can be extended for another year.

Although the mining process helped to understand the current situation of the goods and services acquisition process, the data collected was insufficient to model the current functioning (as is) of the goods and services acquisition process. For this reason, it was necessary to hold meetings with the users seeking to give more quality in the modeling of the process.

In order to obtain a complete representation and a final 'as is' model of the process with all details, it was decided to bring together the users of the departments responsible for each activity of the process, resulting in a report with information on the author, flow activities and description activities. This more detailed analysis allowed the development of an 'as is' model of the process which is closer to reality, with 25 activities being identified in the process. With the current model (as is) defined, it was possible to move on to the next stages of application of BPM at the institute.

5 Final Considerations

This research presented the description and analysis of process modeling using data mining at a Federal Institute of Higher Education in Brazil.

The 'as is' modeling was developed using process mining techniques in the database of the information system of the organization, and through interviews and meetings with the process users.

Due to the fact that the nature of the products and services requested in the institute's system is very varied, the processes are not standardized, and the use of the information system is not managed in a disciplined way, data mining was not sufficient enough to identify the complete processes and the flow of activities in detail.

Despite this, data mining facilitated the definition of instance flows and made it possible to detect critical points in the process, such as points with a long duration of time. With data mining, the analysis time, meetings and modeling effort of the analysts and users were reduced.

In organizations that have a high number of processes, such as the case of the organization analyzed, data mining has been demonstrated to be a tool that facilitates and streamlines the process of identifying, modeling and mapping processes in the application of BPM.

It also showed that the efficiency of data mining depends on how the organization uses its information system. As the processing of the purchasing process does not follow a logical sequence of activities or an event log, it was not sufficient for the modeling of this process, requiring the use of interviews and meetings with users.

As this research presented the results of only one case, its conclusions cannot be generalized, for which an investigation of other institutes and/or other types of organizations would be necessary. Also, it is suggested that future research might reanalyze the process of acquiring goods and services at the organization studied after the implementation of the model 'to be', and verify how the process is behaving.

References

1. Smith, H., Fingar, P.: *Business Process Management (BPM): the third wave*. Meghan- Kiffer Press (2007).
2. Barretta, A., Busco, C.: Technologies of government in public sector's networks: in search of cooperation through management control innovations. *Management Accounting Research*, Boston, vol. 22, n. 4, pp. 211–219 (2011).
3. Segatto, M., Pádua, S. I. D., Martinelli, D. P.: Business process management: a systemic approach? *Business Process Management Journal*, Ribeirão Preto, vol. 19, n. 4, pp. 698–714 (2013).
4. Niehaves, B., Plattfaut, R., Becker, J.: Business process management capabilities in local governments: a multi-method study. *Government Information Quarterly*, Munster, vol. 30, n. 3, pp. 217–225 (2013).
5. Pourmirza, S. et al.: A systematic literature review on the architecture of business process management systems. *Information Systems*, Eindhoven, vol. 66, pp. 43–58 (2017).
6. Avelar, C.F.P., Rocha, T.A.H., Cruz, F.J.S.: Data mining: a review of the Administration literature - Mineração de dados: uma revisão da literatura em Administração. *Vianna Sapiens: Juiz de Fora*, vol. 8, n. 2 (2017). Disponível em: <http://www.viannasapiens.com.br/revista/article/view/232/212> last accessed 2020/02/17.
7. Dumas, M. et al. *Fundamentals of business process management*. Springer Berlin, Tartu, p. 25–100, 2013.
8. Recker, J., Mendling, J.: The state of the art of business process management research as published in the bpm conference. *Business and Information Systems Engineering*, vol. 58, n. 1, pp. 55–72 (2016).
9. Aalst, W. M. P. V. D., Larosa, M., Santoro, F. M.: Business process management. *Business and Information Systems Engineering*, vol. 58, n. 1, pp. 1–6 (2016).
10. Morais, R. M. et al.: An analysis of BPM lifecycles: from a literature review to a framework proposal. *Business Process Management Journal*, vol. 20, n. 3, pp. 412–432 (2014).
11. Milton, S. K., Johnson, L. W.: Service blueprinting and BPMN: a comparison. *Managing Service Quality: an international journal*, Melbourne, vol. 22, n. 6, pp. 606–621 (2012).
12. Aalst, W. V. D.: Process mining: the missing link. *Process Mining*, Springer Berlin Heidelberg, Berlin, pp. 25–52 (2016).
13. Recker, J.: Opportunities and constraints: the current struggle with BPMN. *Business Process Management Journal*, vol. 16, n. 1, pp. 181–201 (2010).
14. Alharbi, A.; Bulpitt, A.; Johnson, O.: Improving pattern detection in healthcare process mining using an interval-based event selection method. *Lecture Notes In Business Information Processing*, Barcelona, pp. 88–105 (2017).
15. Tiwari, A., Turner, C.J., Majeed, B.: A review of business process mining: state-of-the-art and future trends. *Business Process Management Journal*, Cranfield, vol. 14, n. 1, pp. 5–22 (2008).
16. Ferreira, D. R.: Event logs: a primer on process mining. Springer International Publishing, Eindhoven, pp. 1–13 (2017).
17. Dustdar, S., Hoffmann, T., AALST, W. V. D.: Mining of ad-hoc business processes with teamlog. *Data and Knowledge Engineering*, Vienna, vol. 55, n. 2, pp. 129–158 (2005).