Transdisciplinary Engineering: Crossing Boundaries
M. Borsato et al. (Eds.)

doi:10.3233/978-1-61499-703-0-1019

Process Mining to Knowledge Discovery in Healthcare Processes

Gustavo RIZ, Eduardo Alves Portela SANTOS, Eduardo de Freitas Rocha LOURES[1]
Pontifical Catholic University of Paraná - PUCPR / PPGEPS

Abstract. Healthcare processes are complex and require a high-level of interdisciplinary cooperation among the different specialists and sectors involved in their delivery. Information flows among organizational entities, sectors, areas and employees represent possible low process interoperability risks as well as non-compliance risks between business rules and actual process deliveries. Besides this complexity, the Brazilian healthcare area has a notorious problem in its public and private health care systems. These problems are of structural, organizational and financial natures, reflecting the low value attributed to quality and to the actual services in recent surveys of Instituto Data Folha and the Brazilian Ministry of Health (*Ministério da Saúde*). This paper intends to propose an adaptation of Process Mining as an ancillary tool in knowledge discovery processes in healthcare in order to contribute to further improving this area in Brazil. In order to accomplish this, a case study was carried out in the Erasto Gaertner Hospital, located in Curitiba – PR, Brazil, a local reference in cancer treatments.

Keywords. Process mining, process mapping, business rules, organizational mining, healthcare.

Introduction

The healthcare area is complex and has some notorious problems of both organizational and administrative natures. These problems can be described as being related to failures in the organization model, low level of compliance between business rules and the actual processes executed and of interoperability among organizational entities, sectors, areas and staff, lack of training and knowledge by process specialists, and other organizational issues with possible impact in the correct execution of the processes. In addition to these issues, healthcare processes require high-levels of interdisciplinary cooperation and coordination among different specialists such as physicians, nurses, attendants and pharmacists. Despite all this expected interaction, it is normal to find areas working in isolation, since, in many cases, the different areas have no contact or knowledge of what is taking place in other areas. Consequently, often there is no awareness of what is happening in the delivery of health care processes for groups of patients with the same diagnosis, and it is not uncommon for groups of patients with the same diagnosis to receive different test and treatment procedures [1]. This can generate risks of low performance, increased costs and also problems of interoperability among systems and hospital areas when working together

[1] Corresponding authors, {gustavo.riz, eduardo.portela, eduardo.loures}@pucpr.br

in the process (interoperate) to ensure effective exchange of proper, correct and sufficient information.

Brazil has a historical problem with respect to the national public healthcare system. The IDSUS (Índice de Desempenho do Sistema Único de Saúde) performance indicator is used by the Brazilian Ministry of Health to collect information on the quality of National Healthcare System – SUS (Sistema Único de Saúde) distributed in 24 indicators, of which 14 are in connection with access to services and 10, to effectiveness of services received. In the 2011 survey, on a scale of 0 to 10, the national mean score was 5.5. Another study from 2015, carried out by Instituto Datafolha on request from the Federal Council of Medicine (CFM - Conselho Federal de Medicina), heard 2069 people, of whom 93% attributed scores of bad or regular to the public and private health care services in Brazil.

These problems and their underpinning complexity make it necessary to continually monitor health care processes with a view to mitigating the risks of compliance violations, medical errors or damages to patients. Kaymak et al (2012) [2] establish that one of the difficulties in health care is the lack of consistent definitions for its processes with problems that may be listed as pertaining to medical treatment or organizational processes. Such processes are specific (peculiar) and generally display the following characteristics:

a) highly dynamic: as a result of the continuing emergence of new drugs, procedures, treatments and diseases;
b) highly complex: as a function of several factors, such as: complexity of the clinical decision-making process, large data volumes, unpredictability of patients and treatments, and other factors;
c) multidisciplinary: given the existence of a number of specialized departments, medical disciplines and health services providers; and
d) Ad-hoc: due to the high degrees of variability, lack of repetitiveness and low predictability for large scale process analysis.

The importance of having a high-level of governance in these processes is based on the characteristics of the healthcare area in Brazil which is highly regulated and under supervision of several regulatory agencies such as the Brazilian Ministry of Health, Brazilian Health Surveillance Agency (ANVISA), Brazilian Agency of Supplemental Health (ANS), Brazilian Labor and Employment Ministry (MTE) and the municipal and state departments of health. Health services establishments that are not fully compliant with regulations are liable to civil or criminal sanctions ranging from warning notices and fines to sanctions ranging from partial to complete loss of their operating license.

In the healthcare area, the use of process mining is a relatively new, and may be used in making the right clinical decision, thereby reducing costs and improving treatment quality. This has become possible due to the changes occurred in recent decades, through which most health services institutions now have operational support from a control or management system in generating logs of the different activities performed. Hospital information systems (HIS) are examples of these systems, and have aroused increasing interest in Brazil in both, public and private health service providers. The Hospital Information System deployed by the Brazilian SUS (SIH / SUS) [3] is an example of the health services management data systems available. There is also great interest in administrative software applicable in tracking the entire patient service and treatment cycle - from their first contact with reception until the

moment that care and treatment are completed. All this information can provide an accurate perspective of the processes executed.

In this context, Processes Mining is important in that it is capable of, based on the event log information, automatically provide a description of the processes currently applied, and thereby generate the current flow of activities based on the models generated, enabling high-levels of knowledge on the activities performed within processes. In addition, these techniques are applicable to event logs originated by different organizations making it possible for discovery process outcomes to be obtained in relatively short timeframes, and, therefore, deployment in the healthcare area [4].

This paper targets applying process mining through organizational mining in the healthcare processes in order to obtain knowledge on organizational flows, organizational structures and social network analysis among the organizational entities. In this paper, we describe process mining and organizational mining in section 1, section 2 provides a description of the methodology proposed by this research, section 3 contains the case study applied at Erasto Gaertner Hospital and section 4 brings the conclusions.

1. Process mining methodology

The methodology used in this paper starts by considering an overview of the process mining stages comprised of discovery, conformance and extension, and focuses on the discovery of information on the organizational model through the organizational mining techniques.

1.1. Process mining

Process Mining deals with extraction of process delivery knowledge from the respective process execution logs for a wide variety of processes. According to Van der Aalst et al (2012) [5] the goal of process mining is to discover, monitor and improve the actual process delivery through the extraction of event log information and knowledge. Process mining can be used to support the redesign and diagnosis phases through analysis of the information generated in carrying out the process. Event log-based process mining is capable of automatically providing a description of the actual process, generating the activity and information flows. It is applicable to event logs and data logs generated by a wide range of systems, including those applied in healthcare systems. Figure 1 is an adaptation of the traditional process mining model presented by Will M.P. van der Aals (2008) [6] for health care processes.

Process mining has 3 (three) practical applications: 1) discovery of organizational models with a view to constructing models that reflect current situations; 2) measuring the level of compliance between the actual process deliveries and their business rules; and 3) extension which aims at enriching the actual process model by projecting the information extracted from process logs [5].

These techniques seek to achieve different types of models for different perspectives, such as the process perspective (or activity streams), the perspective of the organization and the perspective of the data generated. The type of model generated depends on the technique used. For each technique, there are specific plugins that may be used through the use of a process mining tool. An example of one of these tools is

the Process Mining Workbench (ProM) an open source framework for process mining algorithms.

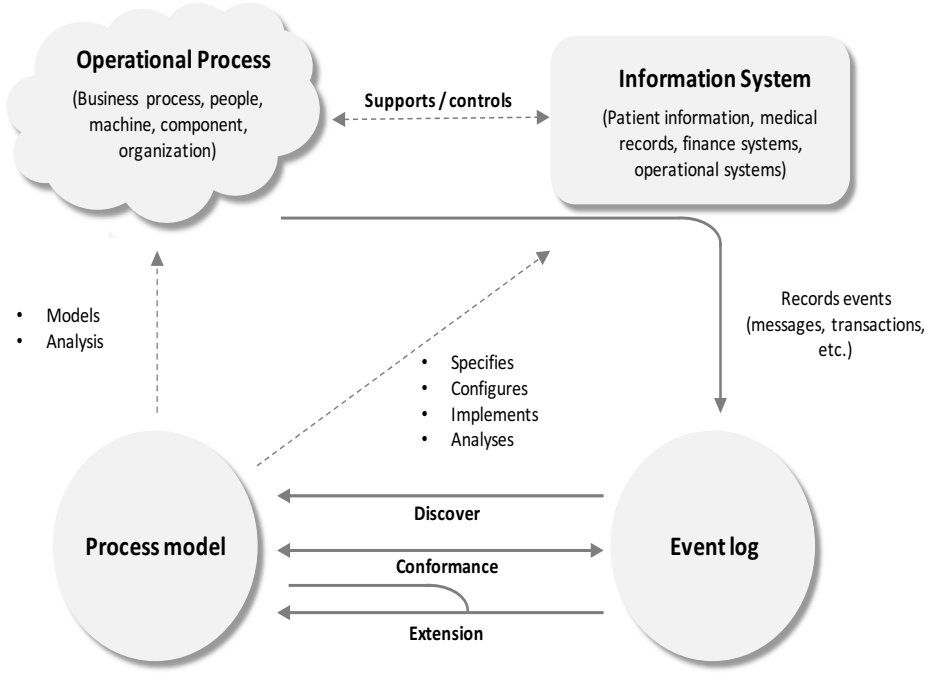

Figure 1. An adaptation of Process Mining overview (Will M.P. van der Aalst et al., 2008) [6].

1.2. Organizational mining

Organizational mining is defined as part of the discovery stage in process mining, responsible for driving understanding of the relationships between organizational entities. There are 3 (three) types of organizational mining: i) organizational model mining; ii) social network analysis; and iii) information flows between organizational entities [6].

Organizational mining is capable of grouping organizational units or rules and discovering the relationships between them. Will M.P. van der Aalst et al (2008) [6] describe that organizational model mining aims at deriving the organizational model from its process logs. By applying process logs, all the elements and interactions comprising the current process can be transcribed.

One of this transcriptions may be the social network analysis. In the ProM Tool, the Social Network plugin provides 5 (five) kinds of metrics to generate social networks - handover of work, subcontracting, working together, similar task and reassignment.

2. Proposed methodology

The use of organizational mining provides a wide range of information to inform knowledge discovery in health care processes, and can generate knowledge about the hospital's processes, such as the interactions between organizational entities and hospital staff, what types of information are exchanged and the frequency in which these information items are exchanged.

As described in Figure 2, there are 3 (three) types of organizational mining. We propose to use two of them: (a) Organizational model mining; and (b) Social Network Analysis. Information flows between organizational entities is out of scope because is used only to discover the information exchanged between entities. The Social Network Miner and Organizational Miner plug-ins in the ProM tool can be used to generate this analysis. The information obtained in this health care process knowledge discovery process will be transcribed into a comparison matrix to display the interactions discovered.

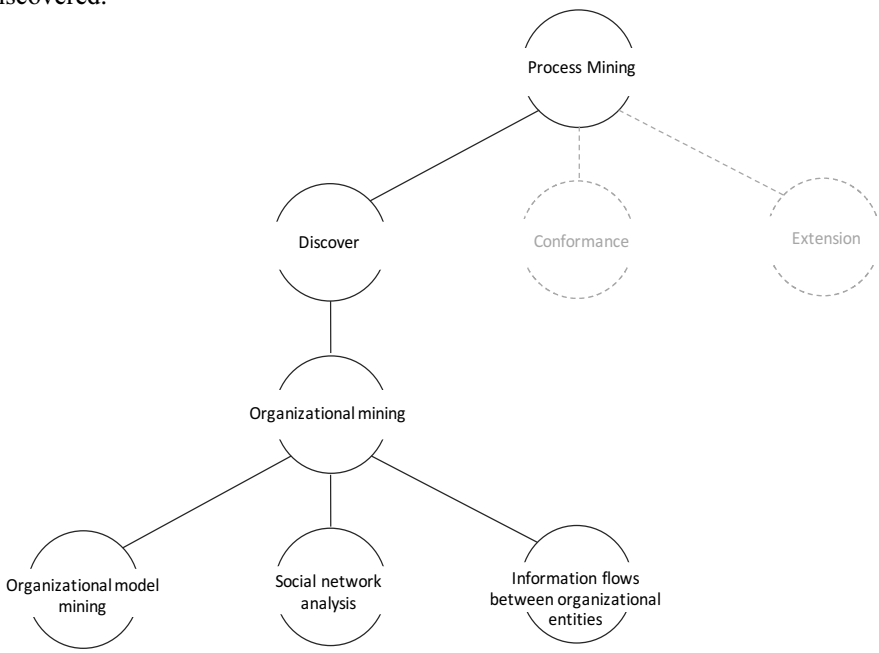

Figure 2. Representation of Organizational Mining.

Pedro Espadinha-Cruz et al (2015) [7] show the use a comparison matrix to demonstrate interactions between two tier suppliers in an automotive supply chain. This matrix represents the internal and external interactions among organizational entities. This information supports the interoperability analysis and may be generated by process mining. Healthcare processes comprise many interactions among different areas, professionals and sectors that can be represented in a comparison matrix (Figure 3).

		RELATIONSHIP AMONG ENTITIES																					
		Area 1						Area 2						Area 3						Area 4					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Area 1	1																								
	2																				X				
	3								X																
	4																								
	5							X															X		
	6										X														

Figure 3. Example of a comparison Matrix.

"X" represents the organizational entities' internal and external interactions, but does not show the power of these interactions nor the order in which they occur, for example, whether the interaction is from "A" to "B" or from "B" to "A" for an interaction between organizational entities "A "and" B ". One proposal is to replace "X" using the symbols "→", "←" and "←→" to represent the type of information exchange that occurred in the process between the two organizational entities. This analysis can provide information to support interoperability analysis, define the critical points to be monitored and the most susceptible areas and activities liable to a failure in interoperability.

3. Application of a case study at Erasto Gaertner Hospital

Erasto Gaertner Hospital (HEG) is located in Curitiba, Brazil, specializes cancer treatment. The choice of HEG is due to the large volume of patients treated. In accordance with the HEG numbers, 310,895 patients were addressed in 2015, generating 1,364,532 patient care procedures.

For this survey, the process selected is chemotherapy treatment. Since the focus is on the patient, the goal is to analyze the path taken by the patient in following the hospital processes, from their first contact with the hospital until the completion of the treatment.

The process support system is Philips Tasy, which was not designed to reflect process dynamics, but was designed for deployment as a data store. Thus, it was necessary to ask the hospital technology team to raise the data needed to draw the path taken by the patient in the process, as described in Table 1.

For this study, a sample of 67 (sixty-seven) patients was requested with the respective database being extracted from Philips Tasy. The data can be then read using a primary key perspective whereby a key enables discovery of the activity flows covered by the patient in the process. Then the extracted base format had to be converted from XLS to XES which to achieve compatibility with the ProM tool. The XES format is an XML-based standard for event logs and can present a wide range of information, including the format for each data item displayed as a timestamp, strings and numeric information.

In the PROM tool, a wide range of information is generated using the log imported, such as the number of processes, cases, events, event classes, event types and originators. PROM supports a wide variety of plug-ins and mining algorithms, such as the Compliance Checker, Organizational Miner and Social Network Miner. The Compliance Checker allows a fitness index between control process model and actual

process model to be obtained, perform a structural analysis of the process executed measuring the amount and consumption, production losses, and process instances that, for some reason, were not completed. The Social Network Miner and Organizational Miner provide information enabling discovery of the interactions occurring between organizational entities. All this information provides an adequate level of knowledge about the relationship between areas, activities, resources and data shared within the process.

Table 1. Data extracted from Tasy System to compose the process log.

Field	Type	Description
Patient ID	Numeric	Patient identification number in the Tasy system
Medical Records ID	Numeric	Medical record number linked to Patient ID
Age	Numeric	Patient age
Sex	String	Male or Female
Treatment type	String	Public (SUS) or Private (health insurance)
Registration date	Date/Time	Date that the register was included in the system
User ID	Numeric	User ID number used to register the information in the system
User Profile	String	User profile linked to User ID used to input user information in the system
Medical specialty	Literal	Medical specialty that originated the service/treatment received by the patient
Date of the activity	Date/Time	Date on which the activity was performed
Activity	String	Description of the activity performed (reception service, medical care, chemotherapy treatment, etc.)
Activity specialist	String	Description of the specialists involved in the execution of the activity

The Organizational Miner identified 37 (thirty-seven) interaction areas related to 11 (eleven) clinical care procedures. The interactions that occur are represented by a connection between the areas and medical specialties enabling the interactions occurring within the process to be established.

In the Social Network miner, several techniques can be applied in analyzing the social networks in the process. Van der Aalst et al (2005) [8] say that social network analysis refers to the collection of methods, techniques and tools in socio-metrics targeted at analyzing the social networks. in order to analyze the relationship among organizational entities, social network analyses were applied, as described in Figure 4.

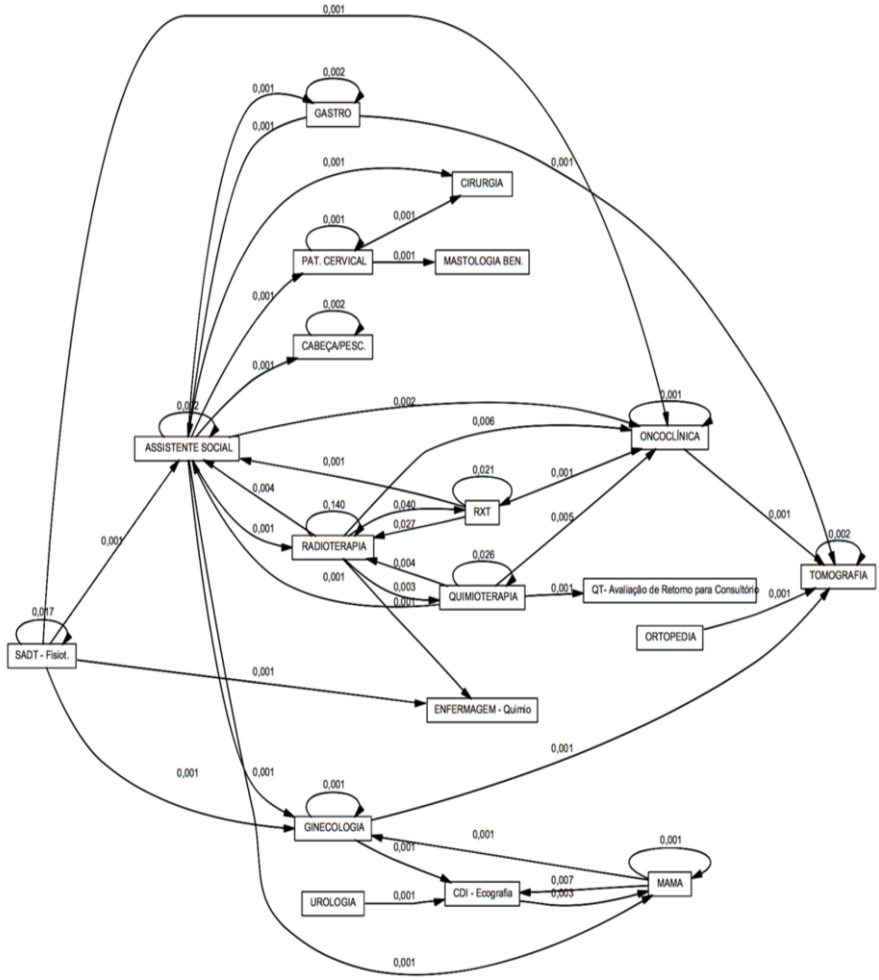

Figure 4. Social graph drafted based on social Network analysis of the event log.

By applying the organizational mining plug-in, information about the organizational structure, organizational entities carrying out similar tasks and organizational entities working together was obtained. An example of organizational mining results is described in Figure 5. Based on this information, a wide range of interaction maps can be generated and display, for instance, the interaction between organizational entities shown in Figure 6.

All the interactions occurring in the process were mapped by applying a comparison matrix. This led to the discovery that the areas with more interactions are Caseworkers (12 interactions), Oncology Clinic (7 interactions), Gynecology and Radiotherapy (6 interactions each). These areas are more likely to have process failures, given the high level of interoperability involved. Poor interoperability among these areas in the exchange of information could adversely impact the entire process.

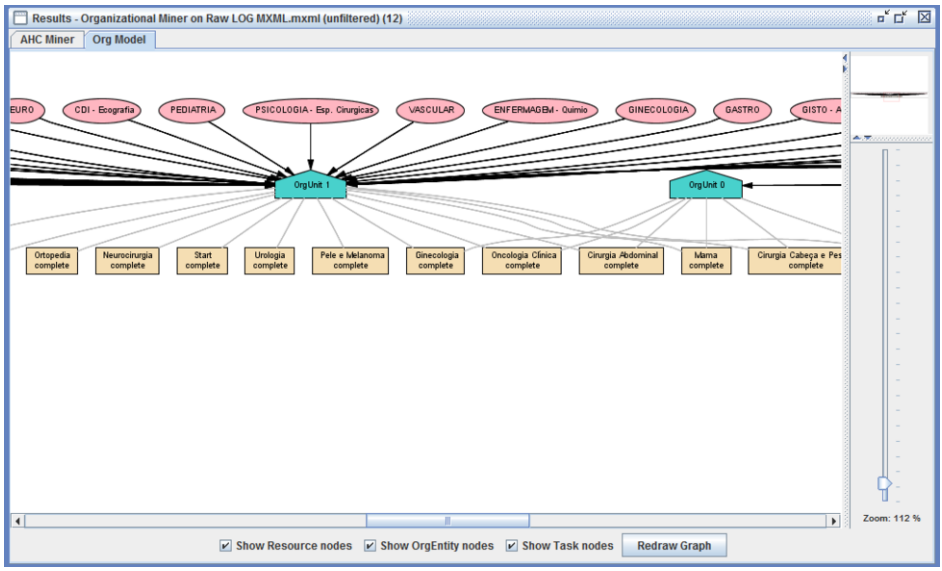

Figure 5. Organizational mining result.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	# Interactions
ASSISTENTE SOCIAL	1	↔	→		↔		↔	→	→		→		→		←	↔	←	←			12
CABEÇA/PESCOÇO	2	←	↔																		2
CDI - ECOGRAFIA	3							←	↔											←	3
CIRURGIA	4	↔											←								2
ENFERMAGEM	5															←		←			2
GASTRO	6	↔					↔												→		3
GINECOLOGIA	7	←		→				↔	←									←	→		6
MAMA	8	←		↔				→	↔												4
MASTOLOGIA	9												←								1
ONCOCLINICA	10	←									↔				←	←	←	←	→		7
ORTOPEDIA	11																		→		1
PAT. CERVICAL	12	←			→					→			↔								4
QT - AVALIAÇÃO DE RETORNO	13														←						1
QUIMIOTERAPIA	14	→									→			→	↔	↔					5
RADIOTERAPIA	15	↔				→					→				↔	→	↔				6
RXT	16	→									→					↔	↔				4
SADT. FISIOT.	17	→				→		→			→							↔			5
TOMOGRAFIA	18						←	←			←	←							↔		5
UROLOGIA	19			→																	1

Figure 6. Comparison Matrix.

4. Conclusions

During The 6 (six) months of fieldwork in HEG, the adequacy of process mining in the healthcare area was tested. The use of process mining is possible due to the flexibility it provides for a variety of different processes. The results generated can be used to inform decision-making and solve problems identified in the process.

In this paper, organizational mining was deployed within the context of healthcare processes. It was possible to map all interactions occurring in the process, and define areas with the highest numbers of interactions. These key areas, which relate to a larger number of areas, are more susceptible to failures in interoperability, making it necessary to monitor processes in order to prevent failures in delivery. The Caseworker

area interacts with 66% of all areas involved in the process, followed by Oncology Clinic (39%), Gynecology (33%) and Radiotherapy 33(%).

The next goal of the research is to extend this model to other processes in different hospitals focusing on obtaining a larger sample and testing the strengths and weaknesses of process mining in healthcare applications, discovering and measuring the information that is exchanged between areas and mature this methodological approach to other processes and hospitals, in order to improve the service and quality of medical and hospital care delivered in both, public and private hospitals, reducing costs and improving revenue.

References

[1] R.S. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst and P.J.M. Rakker, Process Mining in Healthcare – A Case Study in Healthcare – A Case Study in a Dutch Hospital, in Biomedical Engineering Systems and Technologies – Communication in Computer and Information Science, 25, 2009, pp 425-438.

[2] U. Kaymak, R. Mans, T. van de Steeg, M. Dierks, On Process Mining in Health Care, In: *IEEE International Conference on Systems, Man, and Cybernetics*, October 14-17, Seoul, Korea, 2012.

[3] S. Azevedo Bittencourt, L.A. Bastos Camacho, M. do Carmo Leal, O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva – Cad, *Saúde Pública*, Vol.22, 2006, No.1, Rio de Janeiro.

[4] A.J.S. Rebuge, *Business Process Analysis in Healthcare Environments*, Master thesis, The Technical University of Lisboa, 2012.

[5] W. Van der Aalst, Process mining: Overview and opportunities, In: *ACM Transactions of Management Information Systems (TMIS)*, 3(2012), No. 2: 7.

[6] M. Song and W.M.P. van der Aalst, Towards comprehensive support for organization mining, *Decision Support Systems*, Vol. 46(1), 2008, pp. 300-317.

[7] P. Espadinha-Cruz, A. Gonçalves-Coelho, A. Mourão, A. Grilo, Re-design of an interoperable buyer-seller automotive relationship aided by computer simulation, *In 9th International Conference on Axiomatic Design – ICAD*, 2015.

[8] W.M.P. van der Aalst, H.A. Reijers, M.Song, Discovering Social Networks from Event Logs, *Computer Supported Cooperative Work (CSCW)*, December 2005, Volume 14, Issue 6, pp. 549-593.