IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 4, 2020, accepted October 9, 2020, date of publication October 12, 2020, date of current version October 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030318

Supporting Governance in Healthcare Through Process Mining: A Case Study

SIMONE AGOSTINELLI[1], FEDERICO COVINO[2], GIAMPAOLO D'AGNESE[3], CARMELA DE CREA[3], FRANCESCO LEOTTA[1], AND ANDREA MARRELLA[1]

[1]Department of Computer, Control, and Management Engineering, Sapienza University of Rome, 00185 Rome, Italy
[2]Technical Commission "Health Information Systems," Order of Engineers, 00197 Rome, Italy
[3]GVM Care and Research, San Carlo di Nancy Hospital, 00165 Rome, Italy

Corresponding author: Francesco Leotta (leotta@diag.uniroma1.it)

The work of Simone Agostinelli, Francesco Leotta, and Andrea Marrella was supported in part by the H2020 EU Project DESTINI under Grant 857420.

ABSTRACT Healthcare organizations are under increasing pressure to improve productivity, gain competitive advantage and reduce costs. In many cases, despite management already gained some kind of qualitative intuition about inefficiencies and possible bottlenecks related to the enactment of patients' careflows, it does not have the right tools to extract knowledge from available data and make decisions based on a quantitative analysis. To tackle this issue, starting from a real case study conducted in San Carlo di Nancy hospital in Rome (Italy), this article presents the results of a process mining project in the healthcare domain. Process mining techniques are here used to infer meaningful knowledge about the patient careflows from raw event logs consisting of clinical data stored by the hospital information systems. These event logs are analyzed using the ProM framework from three different perspectives: the control flow perspective, the organizational perspective and the performance perspective. The results on the proposed case study show that process mining provided useful insights for the governance of the hospital. In particular, we were able to provide answers to the management of the hospital concerning the value of last investments, and the temporal distribution of abandonments from emergency room and exams without reservation.

INDEX TERMS Healthcare, process mining, ProM.

I. INTRODUCTION

Nowadays, hospitals try to streamline their processes to deliver high quality care improving revenues and reducing costs. If, on the one hand, hospitals are subject to more and more pressure to work in the most efficient way possible, on the other hand, a growth in the demand for care is expected, especially due to an increasingly aging population. A complicating factor is that healthcare is characterized by highly complex and extremely variable patient care processes, also referred to as *careflows* [1]. In healthcare organizations, a wide range of careflows with different characteristics and requirements are daily managed and executed. The delivery of complex care may involve several departments and organizations, and requires an active collaboration between different professionals and practitioners having heterogeneous skills. Additionally, traces of executions of careflows are available in information systems in the form of raw system logs, which make it difficult for the management of a hospital to have a quantitative awareness of the inefficiencies and bottlenecks related to the enactment of careflows, preventing it to transform its qualitative intuitions into governance countermeasures. As a consequence, healthcare is recognized as one of the most promising, yet challenging, domains for the adoption of process-oriented solutions [2]–[4].

In this direction, we present in this article a process mining project performed in the healthcare domain, specifically in San Carlo di Nancy hospital located in Rome, Italy. The motivation behind the use of process mining techniques relies on the availability of system logs recording the execution traces of the careflows concretely enacted in the hospital. This makes process mining solutions as the best possible candidates to discovering and diagnosing such kinds of healthcare processes. To properly conduct the project, we have followed the well-established PM^2 methodology [5] and employed the open-source process mining framework ProM [6], with the aim to infer meaningful knowledge about the patient careflows from the event logs obtained by the clinical data stored in the hospital information systems.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

Process mining aims at extracting process knowledge from event logs in order to discover, for example, both typical paths followed by particular groups of patients and strong collaboration between different hospitalization wards. We performed several analysis under *(i)* the *control flow* perspective, which focuses on the different activities carried out by patients during the careflows, *(ii)* the *organizational* perspective, which describes the relationship between the hospital resources through social networks, and *(iii)* the *performance* perspective, which looks at the timing perspective of the different activities involved in the patient careflows.

We analyzed three different datasets, corresponding to three different clinical processes (i.e., 'Outpatient Clinic', 'Emergency Room' and 'Hospitalizations'), provided by the hospital in accordance with the privacy regulations. The clinical datasets contain raw data about patients treated in the year 2017. Provided data are the ones required by the Italian National Health Service, which any hospital located in Italy must exchange with the Public Regional Authorities. Thus, the proposed analysis can be potentially repeated nationwide. In addition, it is worth to notice we conducted the analysis leveraging just on the available datasets, without having at hand any pre-defined schema or model describing the patients' careflows.

Results obtained on the proposed case study prove the usefulness of applying process analysis and mining for the governance of healthcare facilities. In particular, managerial staff were able to quantitatively evaluate the results of recent investments, consisting in the construction of a new block of the hospital, which encourage new investments in the next future. Additionally, clues have been obtained, which will be helpful to address the problem of treatment abandonments in the emergency room.

The rest of the paper is structured as follows. Section II provides both an overview of process mining and a description of the ProM framework used to infer process models, social networks and transition systems from event logs. Section III reports previous research works applying process mining and related techniques to healthcare. Section IV describes the different steps we followed in order to conduct a process mining project within a hospital context and obtained results. Finally, Section V provides concluding remarks and threats to validity.

II. PRELIMINARIES

Process mining is a research area that enables decision makers to discover process models from event data, compare expected and actual process behaviours, and enrich/repair models exploiting information retrieved from data (such as bottlenecks, deviations from regulations, etc.). Process mining bridges the gap between traditional model-based process analysis (e.g., process simulation) and data-centric analysis techniques such as machine learning and data mining. In particular, it describes a family of a-posteriori analysis techniques exploiting the information recorded in the *event logs*.

An event log is a collection of *traces*, each representing a *process case* (i.e., a specific execution or instance of a process), with each trace represented as sequence of *events* (related to process activity executions) ordered by timestamp and performed by different *resources* (humans, software agents, robots, etc.) involved in the case. Associated to each event there are some mandatory attributes, such as the activity name and execution time, and other optional attributes such as the cost of executing the activity, the lifecycle phase of the activity (i.e., assigned, started, completed) and the resource performing it.

In our case study, we have analyzed three different care processes, each corresponding to a different event log, namely 'Outpatient Clinic', 'Emergency Room' and 'Hospitalizations'. Each trace in any of the available event logs contains the exact activities performed by a patient and the clinical treatments s/he underwent in the context of a single execution of a care process (i.e., of a careflow instance).

Event logs in process mining are usually stored following the XES (eXtensible Event Stream) IEEE standard [7], which is supported by the vast majority of process mining tools. However, available clinical datasets are not usually stored using this format, thus requiring a preprocessing phase that we will detail in the following section.

Process mining addresses the needs of most system/process owners, which have limited information, or only intuitions, of what are the performance and the bottlenecks of the managed processes. This happens because there is often a significant gap between what is supposed to happen, and what actually happens. Only a concise assessment of reality, which process mining strives to deliver, can help at verifying processes, and ultimately be used in system/process redesign efforts and governance. The idea behind process mining is to discover, monitor and improve real processes by extracting knowledge from event logs.

FIGURE 1 shows how process mining can be applied to the hospital context. A hospital is a physical facility where different resources perform treatments on patients supported by an HIS (Hospital Information System) distributed on different kinds of devices ranging from personal computers, tablets and medical equipment. The work of resources inside the hospital is driven by some kind of implicitly or explicitly defined processes, whose instances are recorded in the databases of the HIS and can be turned into event logs for process mining. Once event logs are available, they can be used as input for the different process mining tasks, which are usually classified in the categories of *discovery*, *conformance checking* and *enhancement*.

Discovery aims at deriving information about the original process model, the organizational context, and execution properties from event logs, without any a-priori model in input. Discovery is done under three different perspectives:

- *control-flow*: given an event log containing a set of traces, a suitable process model describing the behavior seen in the log will be automatically constructed. In particular, the process model describes the causal

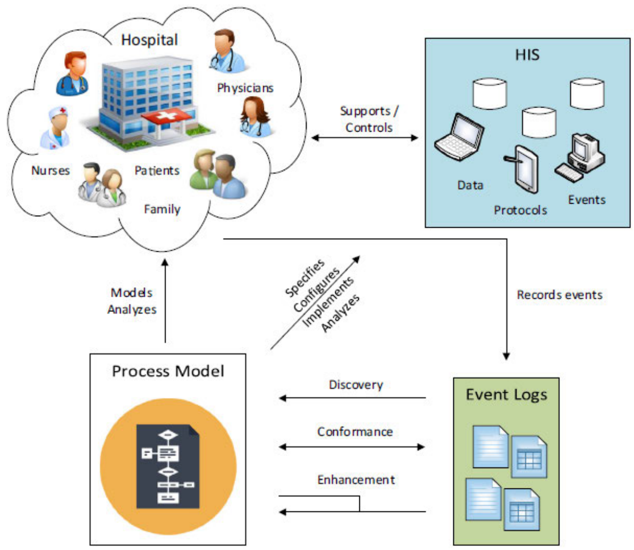


FIGURE 1. Application of process mining to the hospital context, according to [8].

dependencies between activities, focusing on their orderings. The goal is to find a good characterization of all possible paths, expressed in terms of a Petri net or some other flowchart notation (e.g., BPMN, EPC, YAWL, etc.). In this case we talk of *process discovery*. The problem of discovering process models from event logs has been intensely studied in the past two decades, leading to a wide range process discovery algorithms that strike various trade-offs between accuracy, model complexity, and execution [9]. The quality of a discovery process model is expressed through four quality criteria: fitness, precision, generalization and simplicity [10].

- *organizational*: it focuses on information about resources hidden in the log, for instance, on who performed which event. This can give insights in the roles and organizational units or in the relations between individual performers through a social network, that is, which actors (people, systems, roles, and departments) are involved within the system and how are they related [11]–[14]. The goal is to either structure the organization by classifying people in terms of roles and organizational units or to show the social network.
- *performance*: it graphically shows the bottlenecks and all kinds of performance indicators, for example, the average/variance of the total flow time within a process or the sojourn time of a given event.

Given an available process model (discovered or manually defined), *conformance checking* may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations with respect to another process model. Usually, the a-priori model is used as ground truth representing the reality, and it can be employed to check if the discovered model conforms to regulations [15]–[17].

Finally, again in the case a process model is available, *enhancement* techniques allow to enrich the model with the insights extracted from the event logs, for example, either with performance data or with timing information on which bottlenecks are projected [18], [19].

In this contribution, we specifically focus on the application of discovery techniques to allow the management of a hospital to explicitly identify careflows, diagnose problems and set improvement initiatives for governance purposes, by making decisions based on quantitative analysis of performance.

III. RELATED WORK

The application of process mining in the healthcare domain is currently an already tackled field, with several case studies available. Authors in [8] presents a literature review of 74 papers describing several case studies that show the feasibility of process mining techniques in the healthcare domain. For instance, among them, authors in [20] employ some process discovery algorithms implemented in the ProM framework to derive relevant knowledge about a gynecological oncology process in a Dutch hospital In [21], process mining techniques are applied to infer detailed process knowledge from a real case of dental process. The work [22] uses process mining techniques to discover the procedures for treating stroke patients in four different italian hospitals. Here the authors have focused just on the control-flow perspective without considering the organizational and performance perspectives. In addition, Bose and van der Aalst [23] evaluated the applicability of process mining techniques on a real-life event log, taken from a Dutch Academic Hospital, that contains events related to treatment and diagnosis steps for patients diagnosed with cancer.

Some authors conducted their case studies relying on tools different than ProM. For instance, the work [24] presents a case study carried on with Disco.[1] The authors started from a database of the accesses to the services in a local health agency, and produced the activity flow diagram describing these services. In [25], the authors presented a case study formalized with the Declare language [26], with the aim to to show that declarative models are more suitable than imperative ones for describing care processes, which are usually highly unpredictable and unstable. The techniques have been applied in the urology department of the Isala hospital in the Netherlands.

Differently from the previous works, some contributions have introduced new process mining methodologies and methods to be applied on real cases of care processes. For example, in [27], a novel method for providing recommendations in cases of diabetic treatment is implemented. The method combines knowledge management and process mining. The work [28] proposes a novel method to create a similarity metric that is efficient in downgrading the effect of noise and outliers, applied to a real case of a health care

[1] https://fluxicon.com/disco/

institution. The authors of [29] design and develop an interactive visual analytic process discovery tool to explore clinical data from 5784 pediatric asthma emergency department patients. Similarly, the authors of [30] design and implement a Surgical Workflow Management System (SWFMS) that can provide a robust guidance for surgical activities. In [31], a novel event log based approach for extracting valuable medical and organizational information on past executions of the care processes is presented. The work [32] proposes the Clinical Pathway Analysis Method (CPAM) approach that enables the extraction of valuable organizational and medical information on past clinical pathway executions from the event logs of healthcare information systems. The work [33] proposes a data analysis methodology, based on process mining, focusing on the extraction of tangible insights from clinical pathway data by adopting both a drill up and a drill down perspective. In [34], a methodology to perform process diagnostics, based on process mining, was implemented and applied on a case study for a Dutch governmental organization. In [35], a clinical pathway analysis method for extracting valuable medical and organizational information on past diagnosis-treatment cycles is proposed. The method is applied on the clinical pathway processes of a Gynecologic Oncology Department in Belgium. The work [36] proposes a process-oriented methodology for evaluating the impact of IT on two dental processes in Netherlands. Finally, the paper [37] introduces a methodology for the application of process mining techniques that leads to the identification of regular behavior, process variants, and exceptional medical cases.

Similarly to the existing case studies, our work presents a process mining project where data are coming from a non-trivial care process of San Carlo di Nancy hospital in Rome. We conducted the case study following the PM^2 [5] methodology, in order to gain insights about the patients careflows looking at three different perspectives: the control-flow, the organizational and the performance perspective. Even though this is a methodology followed by many of the above mentioned works, each case study has its own distinctive traits, which require a specific approach consisting of selected algorithms and configuration parameters. Additionally, the dataset preprocessing phase is fundamental in order to avoid erroneous deductions due to different kinds of errors in original data.

IV. CONTEXT, METHODOLOGY AND RESULTS

In healthcare organizations, a wide range of care processes with different characteristics and requirements are daily managed and enacted. Specifically, the overall care process related to patient management combines *organizational/administrative* tasks and *clinical procedures*. In the context of our collaboration with a research group of San Carlo di Nancy hospital , we leveraged the PM^2 methodology [5] (cf. FIGURE 2) to rigorously conduct the process mining project. The PM^2 methodology consists of six stages, whose instantiation in the range of the case study will be thoroughly discussed in the next subsections: *(i)* planning, *(ii)* data processing, *(iii)* data extraction, *(iv)* mining & analysis, *(v)* evaluation, and *(vi)* process improvement & support.

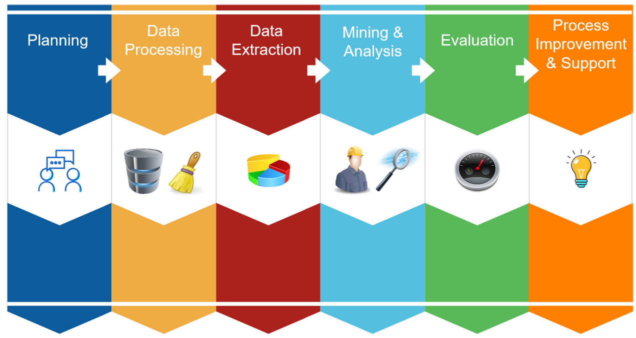


FIGURE 2. An overview of the PM^2 approach.

A. PLANNING

The objective of the planning phase is to set up the project. To this aim, the first activity is to identify management needs. The goal of the hospital was to gain insight about the different careflows related to: *(i)* the procedure for the provision of a healthcare service, *(ii)* the procedure of patients arriving in emergency room, and *(iii)* the hospitalization procedure of patients.

Noteworthy, the hospital management had no previous experience with process mining. To cope whit this, a full-day of process mining lectures was organized to improve the understanding of (care) processes and process mining from the hospital management. A basic understanding of process mining has proved to be beneficial for the hospital management, as it filled the communicative gap, making them able to elicit specific questions and to preside the analysis over the different process mining stages.

B. DATA PROCESSING

The hospital provided us with three clinical datasets:

- *Outpatient clinic*: each row of the dataset stores the information about a single clinic service associated to a patient. A patient with a medical prescription related to a specific healthcare service can go in the corresponding outpatient clinic either with or without the reservation. The total number of records stored within the dataset amounts to 299685.
- *Emergency room*: each row of the dataset represents a single emergency room activity associated to a patient. Once arrived in emergency room, a patient undergoes a first diagnosis and then, depending on the outcome, s/he could be: *(i)* sent home, *(ii)* hospitalized, or *(iii)* transferred to another institute. The total number of records stored within the dataset amounts to 22043.
- *Hospitalizations*: each row of the dataset represents a single hospitalization associated to a patient. Patients may undergo surgery or may receive other types of clinical procedures. The access to hospitalization wards

TABLE 1. Excerpts from the dataset in the intermediate CSV format.

OUTPATIENT CLINIC						
Patient Code	Gender	Birth Date	Request Date	Operation Code	Date Operation	...
M#C_BA3!A=4&5=2M	Male	04/12/1937	04/09/2015	88.73.1	12/01/2016	...
MR_!A_66M48*5=5A	Female	08/08/1966	27/07/2015	95.02	14/01/2016	...
...						
EMERGENCY ROOM						
Patient Code	Gender	Birth Date	Date Entrance	Time Entrance	Issue Code	...
MR#RAI!5&62*5=5F	Female	22/01/1975	31/12/2015	21:44	5	...
BRANRC68_25*5=5M	Male	03/04/1984	01/01/2016	04:42	10	...
...						
HOSPITALIZATIONS						
Patient Code	Gender	Birth Date	Date Check-in	Date Discharge	Date 1st Proc.	...
SN!NMR42A49*5=5A	Female	09/12/1942	02/12/2015	08/01/2016	03/12/2015	...
DCNR#A45P24*5=5V	Male	24/09/1945	18/12/2015	14/01/2016	21/12/2015	...
...						

can take place in different ways. The total number of records stored within the dataset is of 10843 records.

The original datasets contained a sequence of rows consisting of fields whose values were encoded. Additionally, data contained several dirty records consisting of incorrect, missing, imprecise and/or irrelevant data and duplicates. Due to these issues, we converted the datasets from TXT format into and intermediate CSV format in order to filter out the aforementioned issues.

Table 1 shows excerpts of the three intermediate CSV datasets. For each dataset only a small subset of fields is show for the sake of readability. The set of available fields is different for each dataset.

C. DATA EXTRACTION

After cleaning the datasets, the extraction stage aims at extracting event data from the raw data. To achieve such objective, the PM^2 methodology suggests to answer the following questions: *(i)* which is the granularity of the extracted events? *(ii)* which time period should be considered? *(iii)* which data attributes should be extracted? *(iv)* how should data fields be correlated?

This phase, also called *scoping*, is critical and high risk, being it completely manual, tailored on the specific data source and affecting all the downstream phases. In order to build the event logs, there is the need to understand the relevance, the granularity and the variability of each data field of the clinical datasets. Pie charts can be useful to this end, helping the enumeration of all the possible values that a data field can assume. In this way, it is possible to understand to which fields it is possible to associate the concept of event.

Once scoping was completed, three event logs were generated, one for each dataset, by making a correspondence between certain fields and events. Each source dataset row can indeed correspond to multiple events, as clinical activities are denoted by date fields and specific activity codes. Looking at Table 1, examples of such fields can be found.

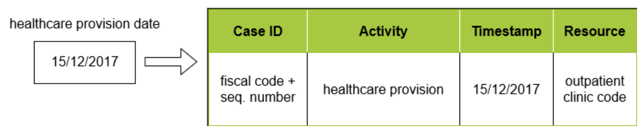


FIGURE 3. Scoping the healthcare provision event.

FIGURE 3 shows an example of how such correspondence between fields and events can be obtained. Starting from the data field *'healthcare provision date'*, it is possible to create the *healthcare provision* activity associated with a case id (the encrypted fiscal code of the patient concatenated with a sequential number), with a timestamp (the value of the raw data field *'healthcare provision date'*), and with the resource that is executing the event. This approach is used for the remaining data fields of the datasets on which was possible to associate the concept of event. Finally, each event is encoded accordingly to the XES standard.

D. MINING & ANALYSIS

In this phase, we applied process mining and process analytics techniques on the three XES event logs obtained through data processing and extraction from the original clinical datasets.

In order to apply process mining, different frameworks and toolkits do exist. In this article, we employed ProM [6], an extensible open-source framework where various process mining algorithms, in the form of plugins, have been implemented. It is platform independent as it is realized in Java. The framework provides easy to use user interface functionality, a variety of model type implementations (Petri nets, BPMN notations, YAWL, etc.), the loading and filtering of event logs, several mining algorithms to be applied and the corresponding visualization of results. Notably, ProM allows the chaining of different plugins to obtain complex result starting from one or more complaint objects (i.e., event logs, process models, etc.). In the majority of cases, the starting input is an simple event log expressed in the XES format.

TABLE 2. List ProM 6.7 plug-ins employed to perform analysis.

Plug-in	Description	Input	Output
Inductive Visual miner [38], [39]	Discovers a model using inductive mining	XES log	PM*
Transition System miner [40]–[42]	Discovers a transition system based on a state representation function and a log	XES log	TS*
Transition System analyzer [40], [41], [43]	Enrich the transition system with timing information	XES log + TS*	TS* enriched with timing
Simple log filter	Filtering a log by using simple heuristics	XES log	XES log
Log visualizer	Allows the visualization and inspection of event logs	XES log	XES log characteristics
Explore event log	Allows a graphical visualization of traces of events	XES log	graphics
Dotted chart analysis [44]	Creates a dotted chart showing all events at glance	XES log	dotted chart
Time based log filter	Filter out traces based on a selected area	XES log	XES log
Social network miner [45], [46]	Creates a social network based on a selected criterion	XES log	SN*

*PM = Process Model
TS = Transition System
SN = Social Network

For the analysis task, we employed ProM version 6.7[2] and Table 2 shows the specific plugins employed for the conducting the experiments discussed in the next sections. In particular:

- Different kinds of filtering and visualization algorithms have been employed to focus the analysis on certain process cases, according to attribute values. In particular, we employed *(i)* the Simple log filter plugin, which filters the XES log by combining a number of simple heuristics that can be configured individually using a wizard, *(ii)* the Time based log filter plugin, which filters out traces according to a specific time range selection, *(iii)* the Explore event log plugin, which allows a graphical visualization of traces of events, e.g., coloring events by event classes or sort traces by length, and *(iv)* the Log visualizer plugin, which allows the visualization and inspection of event logs, e.g., the number of traces and events or the number of resources performing the events.
- Process discovery is performed using the Inductive Visual Miner [38], [39] when the focus is on the process structure. This algorithm produces a graph as output that describes, through a graphical notation, the behaviour of the patients and hospital resources involved in the careflows. The graph can be potentially converted in a BPMN diagram.
- To investigate the conditions that trigger the process moving from one state to another, we employed the Transition System miner [40]–[42]. This plugin produces as an output a transition system that can be then enriched, for example, with the Transition System analyzer [40], [41], [43] to add information about *(i)* how much time is required to execute an activity; *(ii)* how much time is spent to reach an activity from the beginning of the care process and *(iii)* how much time is remaining in order to complete the careflow from a specific activity.
- When the interaction between resources is of interest, the Social network miner [45], [46] has been employed.
- If events must be represented directly, the "Dotted chart analysis" plugin [44] has been used.

Initially, we focused on gaining general insights about the careflows. In particular, different views (versions of the dataset) have been created, by varying the definition of *event*. This is due to a limitation of the XES standard, and in particular in its *organizational extension* [3] with respect to the notion of *resource*. This extension and, as a consequence, all the process mining algorithms currently available in ProM, allow to associate to each event a single resource.

Conversely, in hospitals, there are often different resources involved in each activity and, as a consequence, associated to each event: human resources (of different categories), instrumental resources and infrastructural resources. To cope with this lack, we created different views of the same event data, resulting in different XES event logs, in which the concept of event was associated with the activity performed or with the infrastructural resource involved. This allowed us to analyze the careflows both in terms of the clinical path followed by the patients and in terms of physical movements within the hospital facilities (i.e., emergency room, hospitalization wards, and operating rooms).

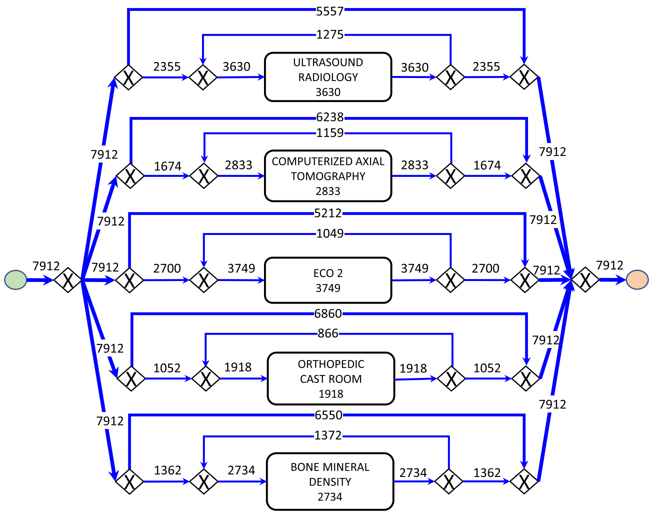
This first analysis has not identified phenomena or aspects unknown to the hospital management. However, it has been possible to quantify a lot of aspects and issues previously known at qualitative level only.

Following these outcomes, the management of San Carlo di Nancy hospital elicited several specific questions about aspects to be further inspected:

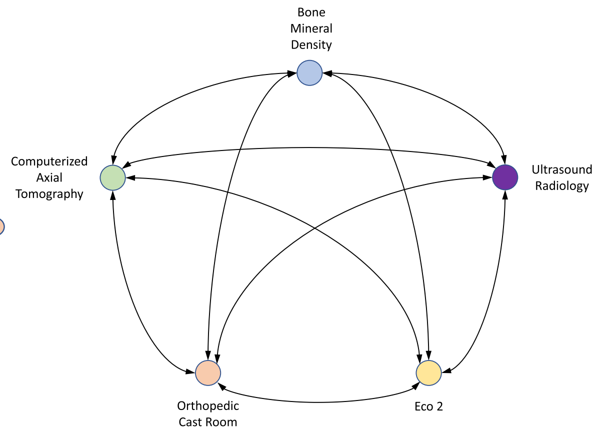
- *Q1: An in-depth analysis of the outpatient clinic services provided by the radiology department. Additionally, the distribution over time of patients without reservation was needed.*
- *Q2: What is the temporal distribution of the patients' abandonments from the emergency room before or in the middle the medical examination?* The goal declared by the hospital management is to adequately adapt its resources to decrease the number of such drop-outs by

[2] http://www.promtools.org/doku.php?id=prom67

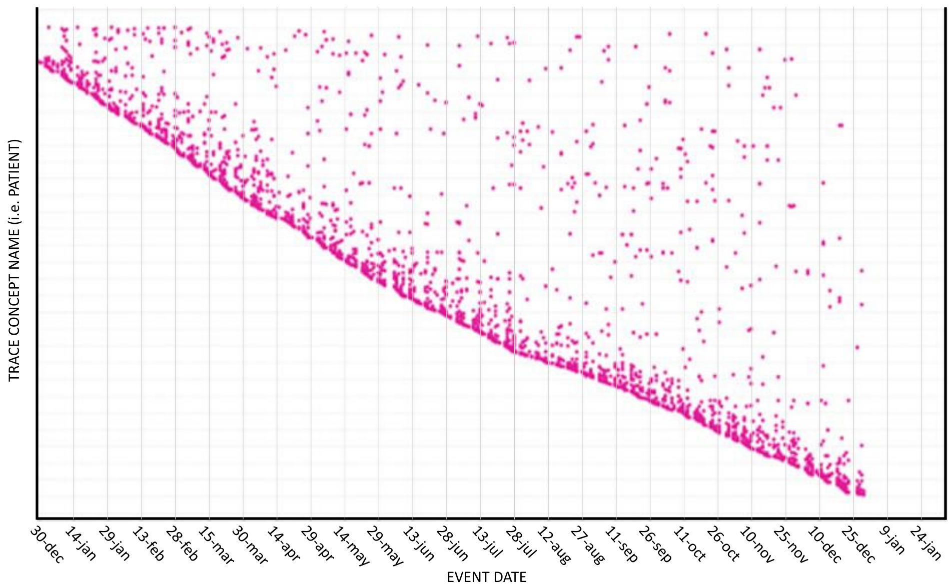
[3] See http://www.xes-standard.org/org.xesext



(a) Patients' distribution among the different radiology services/ sub-departments



(b) Social network interaction of the sub-departments



(c) Distribution of patients without reservation throughout the observation year

FIGURE 4. Outpatient clinic - radiology department.

bringing as many people as possible to the end of medical examinations.

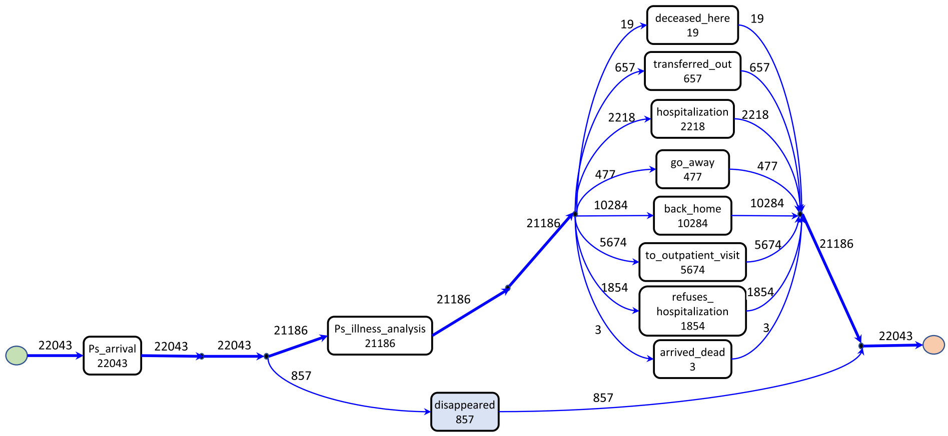
- *Q3: Is the performance of the newly introduced operative block*[4] *better than the pre-existing operative block?* In case a lower average duration of surgical interventions is confirmed, the technological and logistic features of the new block could be replicated in the pre-existing operative block.

1) Q1 - OUTPATIENT CLINIC - RADIOLOGY DEPARTMENT

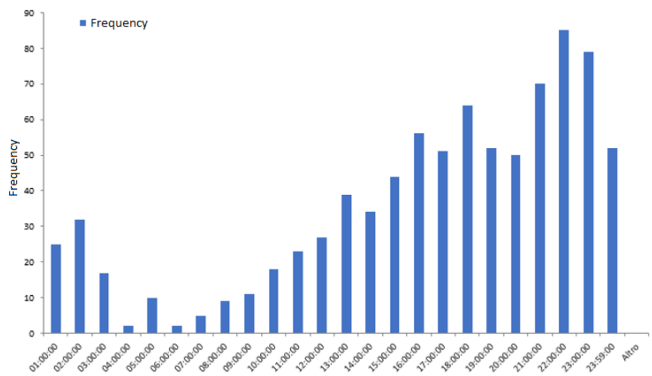
The XES event log derived from the *outpatient clinic* dataset contains activities of three categories, namely *(i)* the emission of the medical prescriptions, *(ii)* the booking of healthcare services, and *(iii)* the execution of these services. Each of the service is performed by a different sub-department. Since the event log contains all the healthcare services of the hospital, we filtered it by considering only those provided by the radiology department thanks to the "Simple log filter" plugin of ProM.

Among the different discovery algorithms in ProM, here the Inductive Visual miner was employed because it produces structured models and it is able to recognize repetitions of activities of single cases. FIGURE 4a shows the resulting process represented as a BPMN diagram, where we only focus on service execution activities. Here, it is possible to see the number of patients involved in each radiology

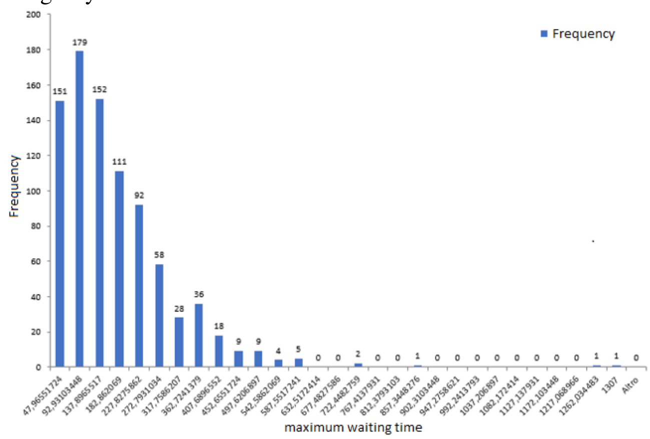
[4] A block is a building of the hospital.



(a) Careflow describing emergency room activities



(b) Distribution of the patients who left the emergency room before the medical examination



(c) Distribution of the maximum waiting time

FIGURE 5. The ***'disappeared'*** **phenomenon.**

sub-department. In the observed period, a total of 7912 people used the services of the radiology laboratory. In particular, 1052 patients passed through the Orthopaedic Cast room, 1362 through BMD - Bone Mineral Density, 1674 through CAT - Computerized Axial Tomography, 2700 through Eco 2, and 2355 through Ultrasound Radiology, with possibly the very same service repeated on the same patient multiple times (e.g., 1049 patients repeated the ECO 2 analysis).

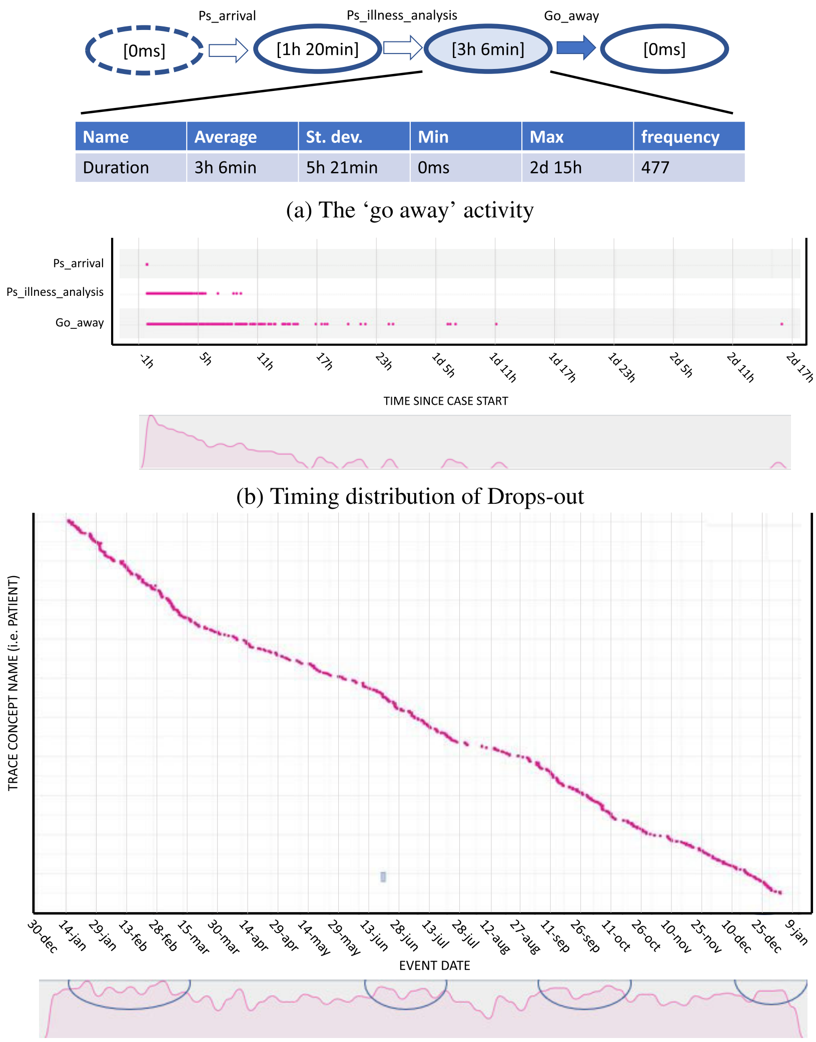
Through social networks analysis (see FIGURE 4b) it emerged that patients can potentially use all resources within the radiology outpatient clinic. The social network shows that the relationships between the resources are many-to-many, which means that the observed dataset describes an interaction between all the possible pairs of resources provided in the dataset.

Of the total 7912 patients, 1398 people did not ask for a reservation. This can be easily detected, by looking at those traces not including a service reservation activity. In order to gain insight about the year distribution of the healthcare services provided without a reservation, within the radiology department, we exploited the "Dotted Chart Analysis" plugin. FIGURE 4c shows on the vertical axis sample case ids, ordered by appearance time, whereas on the horizontal axis the date is shown ranging from December 2015 to December 2016. In the graph, every dot represents a single treated patient. Looking at the distribution of these 1398 people during the observation year, it is possible to see how the level of effort remains constant throughout the year, but with a more pronounced decline in the middle of summer.

2) Q2 - EMERGENCY ROOM - PATIENTS' ABANDONMENTS

With respect to emergency room activities, we focused on the phenomenon of patients leaving the emergency room before or while the medical examination. The XES Event log derived from the *'Emergency room'* clinical dataset contains the following types of activities: *(i)* the arrival of the patient in the emergency room, *(ii)* the medical examination and *(iii)* all the possible outcomes of a medical examination.

FIGURE 5a depicts the careflow mined with the "Inductive Visual miner" plugin (this time no BPMN representation has been extracted). In total, 857 patients leave the emergency room before the medical examination (denoted with the "*disappeared*" activity) and 477 in the middle of it (denoted with the '*go away*' activity).



Name	Average	St. dev.	Min	Max	frequency
Duration	3h 6min	5h 21min	0ms	2d 15h	477

(a) The 'go away' activity

(b) Timing distribution of Drops-out

(c) Frequency distribution of Drops-out

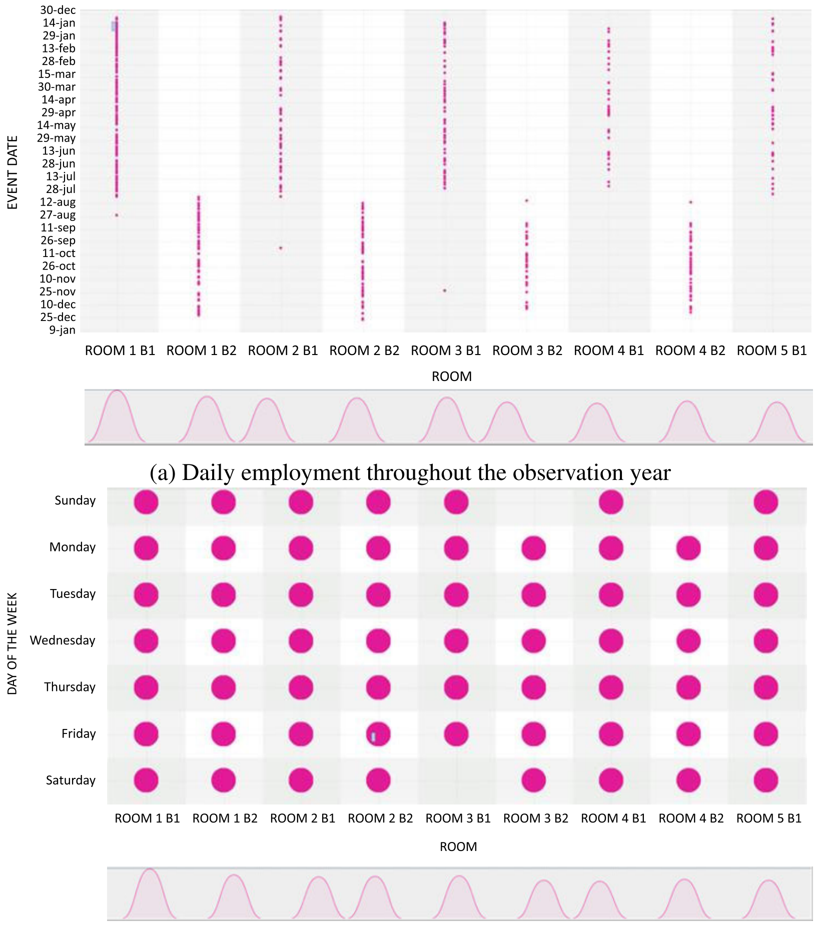
FIGURE 6. The 'go away' phenomenon.

For a further investigation, as requested by the hospital management, we leveraged on process analytic techniques in order to study the hourly distribution of the 857 *'disappeared'* patients. FIGURE 5b shows the number of patients leaving the emergency room before the medical examination, for each hour of the day. Instead, the distribution of the maximum waiting time before leaving the emergency room is depicted in this FIGURE 5c. Looking at the distribution of the maximum waiting time, we had 151 cases of abandonment after a maximum waiting time of 48 minutes, 179 cases of abandonment after a waiting time ranging from 48 minutes to 92 minutes, 152 cases of abandonment after a waiting time that goes from 93 minutes to 138 minutes, and so on and so forth.

Concerning the '*go away*' phenomenon, the number of patients involved are 477. Given a patient, in average s/he left the emergency room after 3 hour and 6 minutes her/his arrival with a standard deviation of 5 hours and 21 minutes (FIGURE 6a). Looking at the dotted chart in FIGURE 6b, on the horizontal axis we have the time and on the vertical axis the activities ("*ps_arrival*", "*ps_illness_analysis*" and "*go_away*"). The picture clearly shows that the medical examination is provided, in most of the cases, before 5 hours since the arrival of the patient to the emergency room. Most drop-outs occur within 10 hours from the access. The peaks of this phenomenon are recorded in January, February, June, September, December as depicted in FIGURE 6c.

3) Q3 - HOSPITALIZATIONS - OPERATIVE BLOCKS

Regarding hospitalizations activities, we focused on obtaining insight on the differences between the new operative block (operating rooms of block B2) compared with the old one (operating rooms of block B1). In particular, we paid attention to the temporal perspective of the surgical interventions.



(a) Daily employment throughout the observation year

(b) Day of the week employment

FIGURE 7. Employment of pperating rooms of blocks B1 and B2.

The XES event log derived from the *'Hospitalization'* dataset contains the following types activities: *(i)* the arrival in emergency room, *(ii)* the medical examination, *(iii)* the hospitalization reservation, *(iv)* the hospitalization, *(v)* the surgical intervention, *(vi)* the hypothetical transfer from a hospitalization ward to another one, and *(vii)* the discharge.

As depicted in FIGURE 7a, the operating rooms of the new block (B2) have been activated in the second half of the year to replace the old operative block (B1). Hence, the 4 rooms of block B2 (room 1, room 2, room 3, room 4) went fully operative at the beginning of August 2017 and the others remained available for exceptional needs. Looking at the employment during the different day of the week (see FIGURE 7b), the operating rooms of the old block (B1) and the new block (B2) are used practically any day of the week, except the room 3 of block B1 never used on Saturday, and both room 3 and room 4 of block B2 are never used on Sunday.

The social network, depicted in FIGURE 8, confirms the separation of these blocks, since there is no interaction between the rooms of block B1 with those of block B2. In addition, it is possible to see that patients performing at least one surgical intervention using several operating rooms.

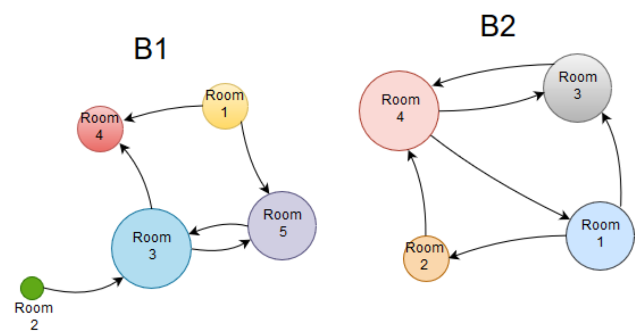
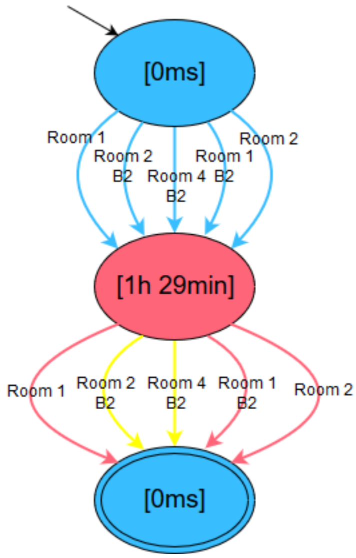
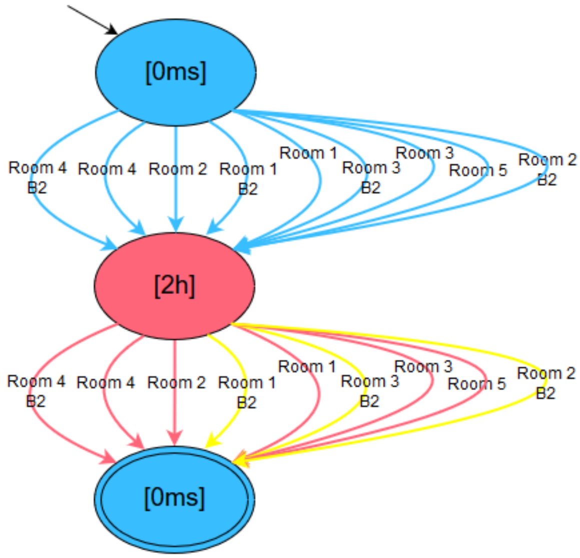


FIGURE 8. Social network representing the interactions of both blocks (B1 and B2) of operating rooms.

FIGURE 9a and FIGURE 9b show the state diagrams for two different kinds of surgical interventions, denoted using the ICD - International Classification of Diseases notation. In particular, the three state diagram has an initial state, correspondent to the phase prior to the assignment of the



(a) Surgical interventions of class ICD 79.35



(b) Surgical interventions of class ICD 06.4

FIGURE 9. Surgical interventions.

room, a second state correspondent to an ongoing surgery, and a final state correspondent to the post-surgery phase. For each state, the average duration is shown. For these two classes, we demonstrated with *ANOVA* that the operative block B2 from the point of view of temporal performance is better than the operative block B1. The yellow color, meaning that the duration of the surgical interventions is lower than the average duration time, is associated only to rooms belonging the the block B2. Even though, as confirmed by ANOVA, in some cases the operative blocks are characterized by different performances, it cannot be argued if this result can be extended to the all classes of surgical interventions.

E. EVALUATION

The objective of the evaluation stage is to review diagrams and results produced during the previous phase and to understand whether they are sufficient to make decisions or if more analysis tasks must be executed. As a consequence, the two phases, mining & analysis and evaluation, can be iteratively repeated until the result is satisfactory for the management. In our case for example, after the first repetition of the mining & analysis phase, the management formulated questions Q1, Q2 and Q3, whose results have been eventually accepted as sufficient for the governance task. Noticeably, the evaluation stage has been simplified, in our case, thanks to the full-day lesson on processes and process mining.

F. PROCESS IMPROVEMENT & SUPPORT

The last step is the process improvement & support stage with the objective to use the gained insights to modify the actual careflows.

During the evaluation stage, the hospital management did not expect the high number of healthcare services provided without reservation. They are currently investigating why this phenomenon is so extensive and if this phenomenon is due to either organizational issues or data errors coming from the information systems. Conversely, they have not been surprised by the results obtained about abandonments of emergency room. The quantitative data will be used to improve this aspect.

Regarding the best performances of the new operative block B2 with respect of the old operative block B1, the hospital management received a confirmation that the investments in more modern instrumental and infrastructural resources have been fruitful.

Additionally, the hospital management is interested in further analysis regarding the economic aspect of the hospital. Indeed, they will provide us with new data to be integrated with the clinical datasets containing the economic values associated to each surgical intervention. The goal of the hospital management is to define the concept of revenue associated to each patient classified by *DRG* - Diagnosis Related Group code. The DRG codes represent the instrument useful to classify the final product of the hospital (i.e., the overall work) for the purpose of remuneration of the hospital activity.

V. THREATS TO VALIDITY AND CONCLUSION

In this article, we conducted a process mining project based on a real case study of an Italian hospital in Rome. In particular, starting from the clinical datasets (containing anonymous records) we were able to discover non-trivial care processes, social networks and transition systems relying on the ProM framework. We gave insights about the process, describing the patient careflows within the hospital, under different points of view by changing the concept of event from time to time. We have also looked both at the organizational perspective of the hospital, describing how the internal resources can communicate among them, and on the performance perspective, analyzing for instance the duration of the surgical interventions.

The presented experimentation is conducted starting from the clinical datasets of the Italian National Health Service that the hospitals interchange with the Public Regional Authorities. As a consequence, the same experimentation can be replicated nationwide establishing a benchmarking between the various healthcare organizations or between different divisions of the same organization. The management of a hospital can easily obtain, thanks to process mining techniques, value-added information that were previously difficult to identify or even undetectable.

The entire process mining project is anyway very sensitive to data preprocessing. Available datasets, as discussed in the paper, are usually not represented in the form of events, i.e., as activities performed by specific resources. Conversely, they are represented in a report form where activities must be argued. In this sense, the governance of the process mining project is of utmost importance, requiring a continuous interaction between data analysts and stakeholders. A powerful tool such as process mining could, in fact, not lead to tangible results for the organization if it is not correctly applied in the organizational areas (and data) in which the root causes of the identified issues persist.

Additionally, clinical datasets contain noise and duplicates that must be removed or at least identified prior to apply process mining. This task may require techniques that are outside the scope of process mining, being instead taken from the wider world of data mining and very similar to what ETL - Extract Transform Load - is for datawarehouses. Thus, to properly apply process mining techniques, some extra effort is required at the outset for pre-processing clinical datasets, but the overhead is compensated by the possibility of automatically discovering and diagnosing healthcare processes.

In our case study, the continuous interaction and collaboration with the management of San Carlo di Nancy hospital allowed us to overcome many of the previously mentioned issues. Still, the lack of certain data (e.g., those related to the costs of activities) did not allow to propose solutions, but only to quantitatively analyze aspects that were already identified, thus allowing the management to make certain decisions. Future work will extend the current analysis by integrating additional data features, that could lead to deeper insights about the performance of the different areas of the hospital.

The main limitation of our study is about *external validity*, i.e., the extent by which the findings can be generalized beyond the scope of the study [47]. The specific healthcare environment in which the evaluation was conducted does not constitute sufficient base to draw general conclusions about the goodness of using process mining in any hospital, since the effectiveness of process mining techniques strongly depends on the quality of event log data recorded by the hospital information systems. Another limitation concerns that the employed datasets keep track of events happened in the span of two months between the end of 2015 and early 2016. For this reason, to confirm the validity and robustness of our findings, we aim to employ the same process mining techniques shown in this study over novel (and more recent) datasets recorded in a longer time frame. Given the above weaknesses, our discussions should be considered as indicative observations rather than conclusive statements.

In conclusion, the lesson learned through the presented case study is that the availability of event data combined with process mining techniques can lead to significant results in terms of process analysis, useful for the hospital management to take final considerations, diagnose problems and set improvement initiatives based on real facts, represented in form of clinical data.

REFERENCES

[1] A. Russo and M. Mecella, "On the evolution of process-oriented approaches for healthcare workflows," *Int. J. Bus. Process Integr. Manage.*, vol. 6, no. 3, pp. 224–246, 2013.

[2] M. Wil and P. van der Aalst, *Process Mining: Data Science in Action*, 2nd ed. Berlin, Germany: Springer, 2016.

[3] F. Cossu, A. Marrella, M. Mecella, A. Russo, G. Bertazzoni, M. Suppa, and F. Grasso, "Improving operational support in hospital wards through vocal interfaces and process-awareness," in *Proc. 25th IEEE Int. Symp. Computer-Based Med. Syst. (CBMS)*, Jun. 2012, pp. 1–6.

[4] T. Catarci, F. Leotta, A. Marrella, M. Mecella, and M. Sharf, "Process-aware enactment of clinical guidelines through multimodal interfaces," *Computers*, vol. 8, no. 3, p. 67, Sep. 2019.

[5] M. L. van Eck, X. Lu, S. J. Leemans, and W. M. Van der Aalst, "PM^2: A process mining project methodology," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Springer, 2015, pp. 297–313.

[6] W. M. V. D. Aalst, B. F. V. Dongen, C. W. Günther, A. Rozinat, E. Verbeek, and T. Weijters, "ProM: The process mining toolkit," *BPM (Demos)*, vol. 489, no. 31, p. 2, 2009.

[7] H. M. W. Verbeek, C. A. M. Joos Buijs, F. Boudewijn van Dongen, and M. P. Wil van der Aalst, *XES, XESame, and ProM 6*. Berlin, Germany: Springer, 2011, pp. 60–75.

[8] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *J. Biomed. Informat.*, vol. 61, pp. 224–236, Jun. 2016.

[9] A. Augusto, R. Conforti, M. Dumas, M. L. Rosa, F. M. Maggi, A. Marrella, M. Mecella, and A. Soo, "Automated discovery of process models from event logs: Review and benchmark," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 4, pp. 686–705, Apr. 2019.

[10] J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van der Aalst, "Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity," *Int. J. Cooperat. Inf. Syst.*, vol. 23, no. 1, Mar. 2014, Art. no. 1440001.

[11] J. Cao, Z. Bu, Y. Wang, H. Yang, J. Jiang, and H.-J. Li, "Detecting prosumer-community groups in smart grids from the multiagent perspective," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 8, pp. 1652–1664, Aug. 2019.

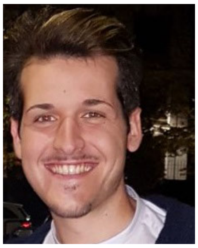
[12] Z. Bu, Y. Wang, H.-J. Li, J. Jiang, Z. Wu, and J. Cao, "Link prediction in temporal networks: Integrating survival analysis and game theory," *Inf. Sci.*, vol. 498, pp. 41–61, Sep. 2019.

[13] Z. Bu, H.-J. Li, C. Zhang, J. Cao, A. Li, and Y. Shi, "Graph K-means based on leader identification, dynamic game, and opinion dynamics," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1348–1361, Jul. 2020.

[14] Y. Wang, Z. Bu, H. Yang, H.-J. Li, and J. Cao, "An effective and scalable overlapping community detection approach: Integrating social identity model and game theory," *Appl. Math. Comput.*, vol. 390, Feb. 2021, Art. no. 125601.

[15] S. J. Leemans, D. Fahland, and W. M. V. D. Aalst, "Exploring processes and deviations," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2014, pp. 304–316.

[16] M. de Leoni, G. Lanciano, and A. Marrella, "Aligning partially-ordered process-execution traces and models using automated planning," in *Proc. 28th Int. Conf. Automated Planning Scheduling (ICAPS)*, 2018, pp. 321–329.
[17] G. de Giacomo, F. M. Maggi, A. Marrella, and S. Sardina, "Computing trace alignment against declarative process models through planning," in *Proc. 26th Int. Conf. Automated Planning Scheduling (ICAPS)*, pp. 367–375, 2016.
[18] D. Fahland and W. M. V. D. Aalst, "Repairing process models to reflect reality," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2012, pp. 229–245.
[19] F. M. Maggi, A. Marrella, G. Capezzuto, and A. A. Cervantes, "Explaining non-compliance of business process models through automated planning," in *Service-Oriented Computing*, C. Pahl, M. Vukovic, J. Yin, and Q. Yu, Eds. Cham, Switzerland: Springer, 2018, pp. 181–197.
[20] R. S. Mans, M. Schonenberg, M. Song, W. M. V. D. Aalst, and P. J. Bakker, "Application of process mining in healthcare–a case study in a dutch hospital," in *Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol.* Berlin, Germany: Springer, 2008, pp. 425–438.
[21] R. Mans, H. Reijers, M. van Genuchten, and D. Wismeijer, "Mining processes in dentistry," in *Proc. 2nd ACM SIGHIT Symp. Int. Health Informat. (IHI)*, 2012, pp. 379–388.
[22] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, and W. V. D. Aalst, "Process mining techniques: An application to stroke care," *Stud. health Technol. Informat.*, vol. 136, p. 573, Dec. 2008.
[23] R. J. C. Bose and W. M. V. D. Aalst, "Analysis of patient treatment procedures," in *Proc. Bus. Process Manage. Workshops*, vol. 99, 2011, pp. 165–166.
[24] D. Antonelli and G. Bruno, "Application of process mining and semantic structuring towards a lean healthcare network," in *Proc. Work. Conf. Virtual Enterprises*. Berlin, Germany: Springer, 2015, pp. 497–508.
[25] M. Rovani, F. M. Maggi, M. de Leoni, and W. M. P. van der Aalst, "Declarative process mining in healthcare," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9236–9251, Dec. 2015.
[26] W. M. P. van der Aalst, M. Pesic, and H. Schonenberg, "Declarative workflows: Balancing between flexibility and support," *Comput. Sci. - Res. Develop.*, vol. 23, no. 2, pp. 99–113, May 2009.
[27] R. D. Dewandono, R. Fauzan, R. Sarno, and M. Sidiq, "Ontology and process mining for diabetic medical treatment sequencing," in *Proc. 7th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, 2013, pp. 171–178.
[28] P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas, and N. Matsatsinis, "Supporting healthcare management decisions via robust clustering of event logs," *Knowl.-Based Syst.*, vol. 84, pp. 203–213, Aug. 2015.
[29] R. C. Basole, M. L. Braunstein, V. Kumar, H. Park, M. Kahng, D. H. Chau, A. Tamersoy, D. A. Hirsh, N. Serban, J. Bost, B. Lesnick, B. L. Schissel, and M. Thompson, "Understanding variations in pediatric asthma care processes in the emergency department using visual analytics," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. 2, pp. 318–323, Mar. 2015.
[30] P. Liebmann, P. Wiedemann, J. Meixensberger, and T. Neumuth, "Surgical workflow management schemata for cataract procedures," *Methods Inf. Med.*, vol. 51, no. 5, pp. 371–382, 2012.
[31] F. Caron, J. Vanthienen, K. Vanhaecht, E. V. Limbergen, J. De Weerdt, and B. Baesens, "Monitoring care processes in the gynecologic oncology department," *Comput. Biol. Med.*, vol. 44, pp. 88–96, Jan. 2014.
[32] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. Deweerdt, and B. Baesens, "A process mining-based investigation of adverse events in care processes," *Health Inf. Manage. J.*, vol. 43, no. 1, pp. 16–25, Mar. 2014.
[33] J. D. Weerdt, F. Caron, J. Vanthienen, and B. Baesens, "Getting a grasp on clinical pathway data: An approach based on process mining," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2012, pp. 22–35.
[34] M. Bozkaya, J. Gabriels, and J. M. van der Werf, "Process diagnostics: A method based on process mining," in *Proc. Int. Conf. Inf., Process, Knowl. Manage.* Berlin, Germany: Springer, Feb. 2009, pp. 22–27.
[35] F. Caron, J. Vanthienen, and B. Baesens, "Healthcare analytics: Examining the diagnosis–treatment cycle," *Procedia Technol.*, vol. 9, pp. 996–1004, 2013.
[36] R. Mans, H. Reijers, D. Wismeijer, and M. van Genuchten, "A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare," *Inf. Syst.*, vol. 38, no. 8, pp. 1097–1115, Nov. 2013.
[37] Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Inf. Syst.*, vol. 37, no. 2, pp. 99–116, Apr. 2012.
[38] S. J. Leemans, D. Fahland, and W. M. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2013, pp. 66–78.
[39] S. J. Leemans, D. Fahland, and W. M. V. D. Aalst, "Using life cycle information in process discovery," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2016, pp. 204–217.
[40] W. M. P. van der Aalst, M. H. Schonenberg, and M. Song, "Time prediction based on process mining," *Inf. Syst.*, vol. 36, no. 2, pp. 450–475, Apr. 2011.
[41] H. E. Verbeek, "BPI Challenge 2012: The transition system case," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2012, pp. 225–226.
[42] W. M. P. Van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. van Dongen, E. Kindler, and C. W. Günther, "Process mining: A two-step approach to balance between underfitting and overfitting," *Softw. Syst. Model.*, vol. 9, no. 1, p. 87, 2010.
[43] W. M. P. Van der Aalst, M. Pesic, and M. Song, "Beyond process mining: From the past to present and future," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Springer, 2010, pp. 38–52.
[44] M. Song and W. M. V. D. Aalst, "Supporting process mining by showing events at a glance," in *Proc. 17th Annu. Workshop Inf. Technol. Syst. (WITS)*, 2007, pp. 139–145.
[45] W. M. P. van der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," *Comput. Supported Cooperat. Work (CSCW)*, vol. 14, no. 6, pp. 549–593, Dec. 2005.
[46] W. M. P. Van der Aalst and M. Song, "Mining social networks: Uncovering interaction patterns in business processes," in *Proc. Int. Conf. Bus. Process Manage.* Berlin, Germany: Springer, 2004, pp. 244–260.
[47] P. Runeson, M. Host, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*. Hoboken, NJ, USA: Wiley, 2012.



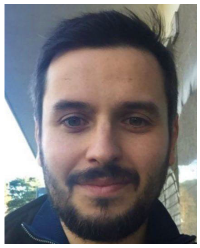
SIMONE AGOSTINELLI received the B.S. and M.S. degrees in engineering in computer science from the Università di Roma. He is currently pursuing the Ph.D. degree in engineering in computer science with the Sapienza University of Rome. His research interests include synthesizing strategies for robotic process automation via process mining and automated planning techniques. In 2019, he received the Forum Award at the 31st International Conference on Advanced Information Systems Engineering (CAiSE).

FEDERICO COVINO received the M.Sc. degree in engineering in computer science from the Sapienza Università di Roma. After a period as a Research Fellow, he started a career in the business process management and enterprise risk management fields for air control traffic and healthcare sectors. TOGAF, COBIT, OCEB2, and Lean Six Sigma certified, he is currently a member of the Technical Commission "Health Information Systems," Order of Engineers, Rome, and a member of the National Association of Technical Officers of the Italian Army. He is the author of scientific and educational articles on mesh-based architectures and e-health. He promotes and coordinates R and S activities in collaboration with Italian Universities and bodies in the field of process improvement, with specific interest in the integration of consolidated methodologies with cutting-edge technologies.

GIAMPAOLO D'AGNESE is currently the Chief Information Officer of the San Carlo di Nancy Hospital, Rome, Italy. He has 20 years of experience in the field of information systems and management control in health, gained in several Italian medium and large hospitals. In particular, he has a wide experience in planning, designing, coordinating, and implementing data processing systems for healthcare, such as laboratory information systems, radiology information systems and PACS, hospital information systems, and systems for managing patients in discharge and transfer to other hospitals.

CARMELA DE CREA received the Ph.D. degree in experimental endocrine, metabolic and surgical sciences from the Department of Surgery, School of Medicine, Catholic University Sacro Cuore, Rome, Italy, in 2003. She has been an Associate Professor of surgery with Catholic University Sacro Cuore, since 2018, and the Chief of the Division of Endocrine Surgery of San Carlo di Nancy Hospital, Rome, since 2016. She has coauthored more than 200 scientific publications. From June 2005 to February 2018, she performed 2707 surgical procedures. From 2013 to 2016, she has been the Chief of the Section of Integrated Diagnostics and Treatment of Thyroid Cancer of the Hospital "A. Gemelli" of Rome. In 2000, she is a Fellow in Endocrine Surgery (a Research Associate Program in Endocrine Surgery) at Rhode Island Hospital, Brown University, Providence, RI. Since 2014, she has been a Visiting Professor of the Section of Endocrine Surgery (Dir. Prof. J. A. Sosa) of Duke University, Durham, NC.

FRANCESCO LEOTTA received the Ph.D. degree in engineering in computer science from Sapienza, in 2014. He is currently an Assistant Professor with the Department of Computer, Control, and Management Engineering, Sapienza Università di Roma, Italy. His research interests include algorithmic, methodological, experimental, and practical aspects in different areas of computer science, including ubiquitous computing, human–computer interaction, and digital humanities. Such topics are challenged in the application domains of smart spaces, smart manufacturing, and cultural heritage.



ANDREA MARRELLA is currently an Assistant Professor with the Sapienza Università di Roma, Italy. His research interest includes how to integrate artificial intelligence with business process management solutions, to untangle complex challenges from the fields of process mining and robotic process automation. He has coauthored more than 70 peer-reviewed publications in renowned international conferences and top journals. Since 2017, he has been the Information Director of the *ACM Journal of Data and Information Quality*.