Chapter 1
Introduction

Abstract Healthcare costs have increased dramatically and the demand for high-quality care will only grow in our aging society. At the same time, more event data are being collected about care processes. Healthcare Information Systems (HIS) have hundreds of tables with patient-related event data. Therefore, it is quite natural to exploit these data to improve care processes while reducing costs. Data science techniques will play a crucial role in this endeavor. Process mining can be used to improve compliance and performance while reducing costs. The chapter sets the scene for process mining in healthcare, thus serving as an introduction to this *SpringerBrief*.

Keywords Healthcare information systems · Process mining · Healthcare · Business process management

Process mining has been applied successfully in a variety of domains, e.g., banking, insurance, logistics, production, e-government, customer relationship management, remote monitoring, and smart diagnostics. Through process mining one can relate the actual behavior of people, machines, and organizations with modeled behavior. This often leads to surprising insights showing that reality is very different from perceptions, opinions, and beliefs stakeholders have. This is particularly relevant for healthcare processes. These processes are often only partly structured with many exceptional behaviors and different stakeholders. Healthcare requires flexibility and ad-hoc decision making. These characteristics make it impossible to apply rigorous Business Process Management (BPM), Workflow Management (WFM), and Business Process Reengineering (BPR) techniques. Clearly, a hospital is not a factory and patients cannot be cured using a conveyor belt system. However, the abundance of data collected in today's hospitals can be used to improve care processes dramatically. Unlike many other domains, there is still room for dramatic improvements in healthcare processes. Process mining can be used to improve compliance and performance while reducing costs. To set the scene, this chapter introduces the application of process mining in healthcare. Section 1.1 discusses the main challenges in healthcare. In Sect. 1.2, process mining is positioned in the broader *data science* context. Subsequently, Sect. 1.3 discusses the application of process mining in healthcare. Section 1.4 concludes the chapter with an outlook on the remainder of this *SpringerBrief*.

R.S. Mans et al., *Process Mining in Healthcare*,
SpringerBriefs in Business Process Management, DOI 10.1007/978-3-319-16071-9_1

1.1 Challenges in Healthcare

Healthcare is facing several challenges. Some of the most urgent challenges become evident when looking at Fig. 1.1. First, at the top of the figure, it is shown that healthcare costs continue to rise. So, there is a need to reduce these costs. Second, the people receiving care are becoming older. This is likely to lead to greater demand for elderly care [1]. Finally, the bottom of the figure shows that the volume of long-term

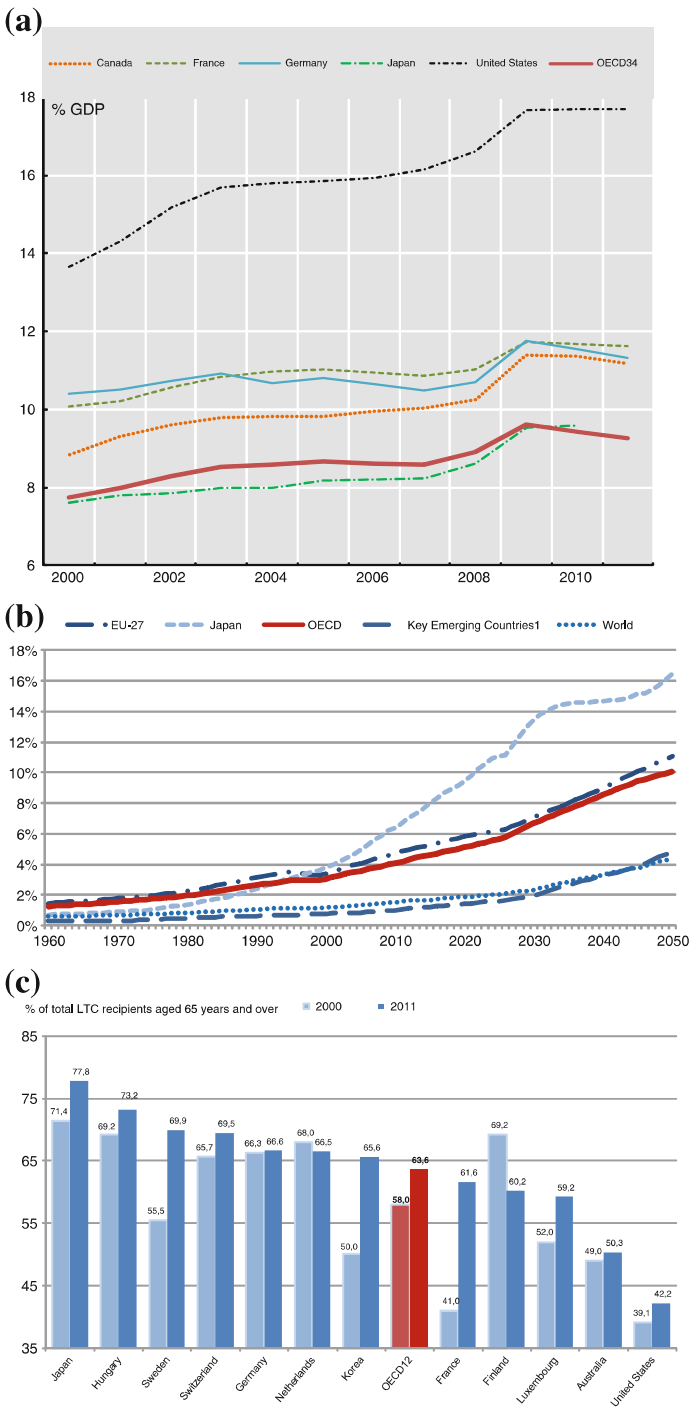

Fig. 1.1 Within healthcare, costs are rising, people are aging, and the demand for care is increasing. **a** Total health expenditure as a share of GDP, 2000–2011. *Source* OECD Health Statistics 2013, http://dx.doi.org/10.1787/health-data-en. **b** Trends in the share of the population aged over 80 years, 1960–2050. *Source* OECD Historical Population Data and Projections Database, 2013. **c** Share of long-term care recipients aged 65 years and over receiving care at home, 2000 and 2011 (or nearest year). *Source* OECD Health Statistics 2013, http://dx.doi.org/10.1787/health-data-en

care increased in the period of 2000 till 2011. For these and the other types of care, further increases are expected. Regarding the care provided, an important health policy issue in many OECD countries relates to long waiting times [2]. These long waiting times cause dissatisfaction as the benefits of treatment are postponed.

The developments shown in Fig. 1.1 illustrate the pressure on today's healthcare organizations. They need to improve productivity and reduce access and waiting times while at the same time reducing costs. One approach to this is to focus on the many complex time-consuming and non-trivial processes that are undertaken within these organizations. Examples of such processes are the preparation and execution of a surgery and the treatment of patients suffering from cancer. In order to give suggestions for improving and redesigning these processes they need to be analyzed. Such an analysis is typically done by conducting interviews. Unfortunately, this is time consuming and costly. Furthermore, typically a *subjective* view is provided on how a process is executed. That is, people involved in the performance of these healthcare processes (e.g., physicians, managers) tend to have an ideal scenario in mind, which in reality is only one of the many scenarios possible. Moreover, in many hospitals "political battles" take place due to organizational issues. Different stakeholders may have different views, e.g., some parties may not be interested in reducing the overall costs and improving transparency. Therefore, in order to give objective suggestions for improving and redesigning processes one needs to exploit the event data readily available. Such an analysis is possible using process mining.

1.2 Process Mining: Data Science in Action

Although our capabilities to store and process data have been increasing exponentially since the 1960-ties, suddenly many organizations realize that survival is not possible without exploiting available data intelligently. This of course also holds for healthcare organizations. Society, organizations, and people are "Always On". Data are collected *about anything*, *at any time*, and *at any place* [3]. Gartner uses the phrase "The Nexus of Forces" to refer to the convergence and mutual reinforcement of four interdependent trends: social, mobile, cloud, and information [4]. The term "Big Data" is often used to refer to the incredible growth of data in recent years. For hospitals of course the goal is *not* to collect more data, but to exploit data *to realize more efficient and effective care processes*.

Obviously, the term "Big Data" has been hyped in recent years. However, there is rapidly growing demand for *data scientists* that can turn data into value. Just like computer science emerged as a new discipline from mathematics when computers became abundantly available, we now see the birth of data science as a new discipline driven by the huge amounts of data available today. Data science aims to use the different data sources to answer questions that can be grouped into the following four categories:

- Reporting: *What happened?*
- Diagnosis: *Why did it happen?*
- Prediction: *What will happen?*
- Recommendation: *What is the best that can happen?*

So, what is a data scientist? Many definitions have been suggested. For example, [5] states "Data scientists are the people who understand how to fish out answers to important business questions from today's tsunami of unstructured information". It is not easy to define the ideal profile of a data scientist. Clearly, data science is multidisciplinary. As Fig. 1.2 shows, data science is more than analytics/statistics. It also involves behavioral/social sciences (e.g., for ethics and understanding human behavior), industrial engineering (e.g., to value data and know about new business models), and visualization. Just like Big Data is more than MapReduce, data science is more than mining. Besides having theoretical knowledge of analysis methods, the data scientist should be creative and able to realize solutions using IT. Moreover, the data scientist should have domain knowledge and able to convey the message well. Figure 1.2 shows a possible profile of the data scientist: different subdisciplines are combined to render an engineer that has quantitative and technical skills, is creative and communicative, and is able to realize end-to-end solutions.

Figure 1.2 deliberately emphasizes the *process* aspect. The goal is not to analyze data, but to improve care processes. *Process mining* aims to *discover, monitor and improve real processes by extracting knowledge from event logs* readily available in today's information systems [6]. Starting point for process mining is an *event log*. Each event in such a log refers to an *activity* (i.e., a well-defined step in some process) and is related to a particular *case* (i.e., a *process instance*). The events

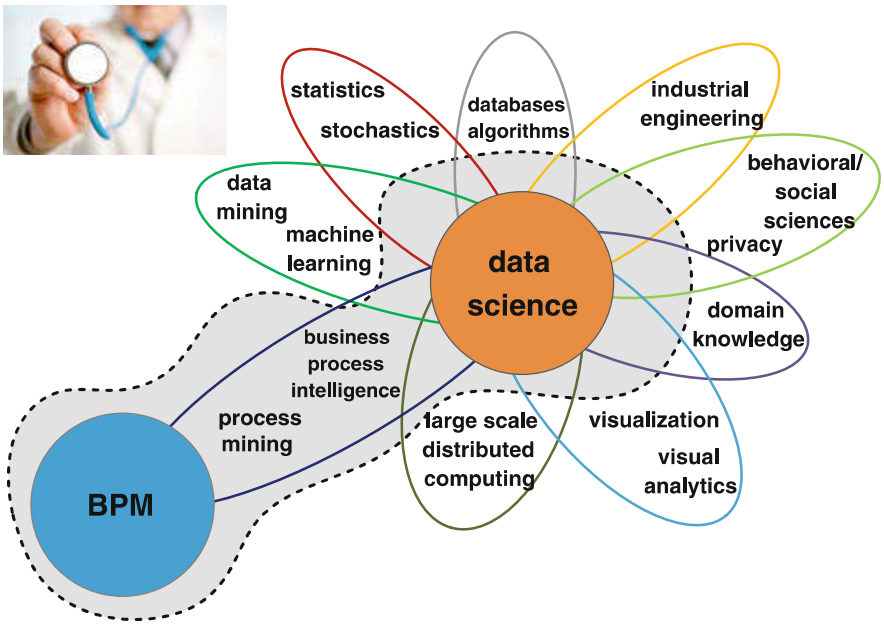

Fig. 1.2 Data science skills that should be combined to realize more efficient and effective care processes

belonging to a case are *ordered* and can be seen as one "run" of the process. Event logs may store additional information about events. In fact, whenever possible, process mining techniques use extra information such as the *resource* (i.e., person or device) executing or initiating the activity, the *timestamp* of the event, or *data elements* recorded with the event (e.g., the age of a patient).

Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining [6]. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand-made or discovered automatically). This technology has become available only recently, but it can be applied to any type of operational processes (organizations and systems).

There are three main types of process mining:

- The first type of process mining is *discovery*. A discovery technique takes an event log and produces a process model without using any a-priori information. An example is the Alpha-algorithm [7] that takes an event log and produces a process model (a Petri net) explaining the behavior recorded in the log.
- The second type of process mining is *conformance*. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa [8].
- The third type of process mining is *enhancement*. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log [6]. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model. An example is the extension of a process model with performance information, e.g., showing bottlenecks.

Process mining techniques can be used in an offline, but also online, setting. The latter is known as *operational support*. An example is the detection of non-conformance at the moment the deviation actually takes place. Another example is time prediction for running cases, i.e., given a partially executed case the remaining processing time is estimated based on historic information of similar cases.

1.3 Applying Process Mining to Healthcare Processes

As mentioned in Sect. 1.1 care organizations are under incredible pressure "to do more for less". To be able to improve processes it is important to understand what is really happening (process discovery) and analyze deviations from the expected or normative process model (conformance checking). Moreover, using the timestamps of events one can identify and diagnose bottlenecks and other inefficiencies (enhancement). Chapter 3 introduces process mining in detail. At this stage it is sufficient to have a rough idea of the results and insights provided by process mining.

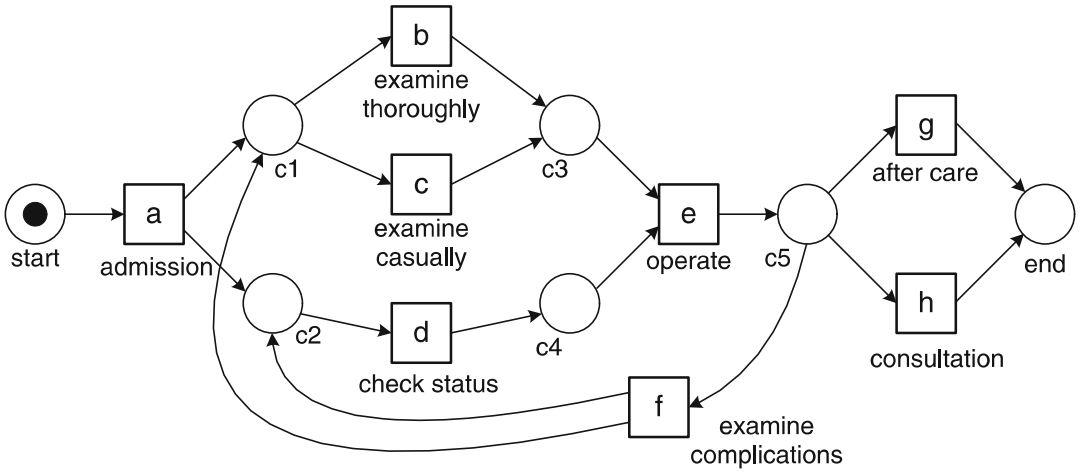

Fig. 1.3 Process model only showing the control-flow. The model is not intended to be realistic and only aims to show the different control-flow constructs in a healthcare setting

Figure 1.3 shows a simplified process model learned from event data. The backbone of the process model is formed by the control-flow, i.e., the ordering of activities. The control-flow is represented in terms of a Petri net, i.e., a bipartite graph of transitions representing activities and places representing states. The process starts by admitting a patient. This activity is modeled by transition *admission*. Each transition is represented by a square. Transitions are connected through places that model possible states of the process. Each place is represented by a circle. In a Petri net a transition is *enabled*, i.e., the corresponding activity can occur, if all input places hold a token. Transition *admission* has only one input place (start) and this place initially contains a token representing a patient that needs treatment. Hence, the corresponding activity is enabled and can occur. This is also referred to as *firing*. When firing, the transition consumes one token from each of its input places and produces one token for each of its output places. Hence, the firing of transition *admission* results in the removal of the token from input place *start* and the production of two tokens: one for output place *c1* and one for output place *c2*. Tokens are shown as black dots. The configuration of tokens over places—in this case the state of the patient's treatment—is referred to as *marking*. Figure 1.3 shows the initial marking consisting of one token in place *start*. The marking after firing transition *admission* has two tokens: one in place *c1* and one in place *c2*. After firing transition *admission*, three transitions are enabled. The token in place *c2* enables transition *check status*. This transition models a review of the medical history of the patient. In parallel, the token in *c1* enables both *examine thoroughly* and *examine casually*. Firing *examine thoroughly* will remove the token from *c1*, thus disabling *examine casually*. Similarly, the concurrence of *examine casually* will disable *examine thoroughly*. In other words, there is an exclusive choice between these two activities. Transition *examine thoroughly* is executed for patients where complications are expected. Less problematic cases only need a casual examination. Firing *check status* does not disable any other transition, i.e., it can occur concurrently with examine thoroughly or examine casually. Transition *operate* is only enabled if both input places contain a token. The

medical history of patient needs to be checked beforehand (token in place *c4*) and the casual or thorough examination should have been completed (token in place *c3*). Hence, the process synchronizes before operating. Transition *operate* consumes two tokens and produces one token for *c5*. Three transitions share *c5* as an input place. This shows that there are three possible scenarios. Etc. The process ends with a token in place *end*.

A process model such as the one shown in Fig. 1.3 can be learned by analyzing events logs describing the activities executed for patients [6]. A possible *trace* for a particular patient is $\langle a, b, d, e, h\rangle$. Note that here we are using short names (e.g., $a = admission$) and do not show the attributes of the various events, e.g., timestamp, resource, and data. Another possible trace is $\langle a, c, d, e, f, d, c, e, f, c, d, e, h\rangle$. An event log can be viewed as a multiset of traces (if we ignored timestamps, etc.). $L = [\langle a, b, d, e, h\rangle^5, \langle a, d, c, e, g\rangle^4, \langle a, c, d, e, f, b, d, e, g\rangle^4, \langle a, d, b, e, h\rangle^3, \langle a, c, d, e, f, d, c, e, f, c, d, e, h\rangle^2, \langle a, c, d, e, g\rangle^2]$ is an event log with 20 cases. Based on this event log most process discovery techniques construct a control-flow model as is shown in Fig. 1.3. This model is indeed able to reproduce the traces observed.

The events belonging to a case are not just ordered. There may be extra information such as the resource (i.e., physician or nurse) executing or initiating the activity, the timestamp of the event, or data elements characterizing the patient. By replaying the event log on the model shown in Fig. 1.3 we can learn additional perspectives and enrich the model as is shown in Fig. 1.4.

As Fig. 1.4 shows, the process model can be extended with additional perspectives: the organizational perspective ("What are the organizational roles and which resources are performing particular activities?"), the case perspective ("Which characteristics of a case influence a particular decision?"), and the time perspective ("Where are the bottlenecks in my process?") [6]. Analysis of the event log shown may reveal that Sara is the only one performing the activities *operate* and *examine complications*. This suggests that there is a "surgeon role" and that Sara is the only one having this role. Activity *examine thoroughly* is performed only by Sue and Sean. This suggests some "physician role" associated to this activity, etc. Techniques for organizational process mining [6, 9] will discover such organizational structures and relate activities to resources through roles. By exploiting resource information in the log, the organizational perspective can be added to the process model. Similarly, information on timestamps and frequencies can be used to add performance related information to the model. Figure 1.4 sketches that it is possible to measure the time that passes between an examination (activities *b* or *c*) and the actual operation (activity *e*). If this time is remarkably long, process mining can be used to identify the problem and discover possible root causes. If the event log contains case-related information, this can be used to further analyze the decision points in the process. For instance, through decision point analysis it may be learned that older patients require multiple operations.

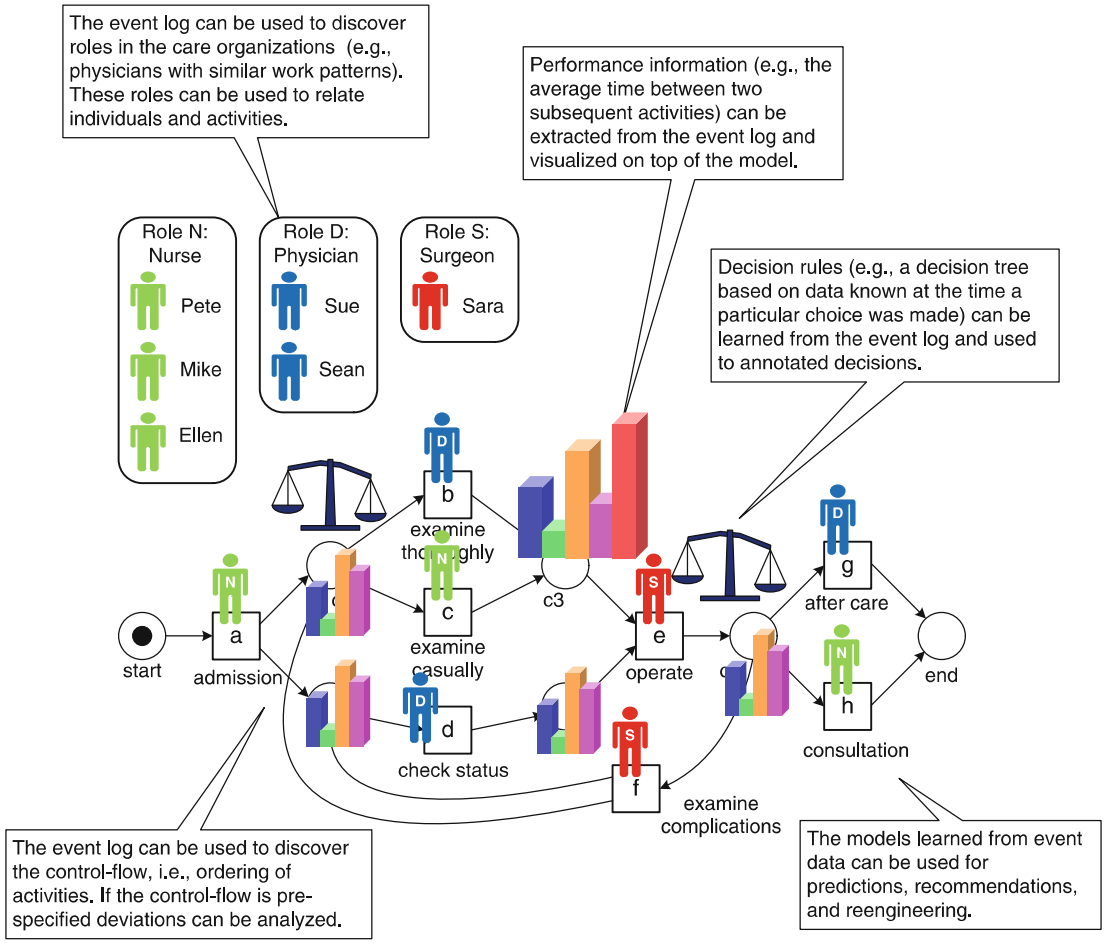

Fig. 1.4 Process mining is not only used to learn the process as it is actually executed: It is also used to understand deviations, to analyze bottlenecks, and to monitor organizational behavior

1.4 Outlook

Process mining [6] aims at extracting process knowledge from so-called *event logs* which may originate from all kinds of systems. Examples of such systems are Hospital Information Systems (HIS) but may also be systems in use at an intensive care storing all diagnostic tests and treatments that have been performed or a laboratory system storing all tests that have been performed on a blood sample. Typically, these event logs contain information about the start/completion of process steps together with related context data (e.g., actors and costs involved). Since process mining uses factual execution data, it allows for obtaining an objective view on how processes are really executed. In this way, there is a clear difference between process mining and more traditional ways of investigating business processes. For example, by conducting interviews there is always the risk that highly subjective information is gathered.

As process mining allows for easily getting insights into the real execution of organizational healthcare processes, not surprisingly, there is a growing uptake of

the technique in the healthcare domain. That is, the obtained insights can be used for example to reduce costs and to improve the efficiency of the care processes. As part of this, the patient satisfaction is expected to grow. Also, in literature, up to now, we have discovered 59 publications in which a real-life application of process mining in healthcare is described (see http://www.healthcare-analytics-process-mining.org/ for an overview). For these applications often only data are taken from one or two systems in order to solve a particular problem. Despite this popularity, *an overview is missing of all the process related data that exists within a HIS*. As a result, it is difficult to *reason about potential applications of process mining within hospitals*.

The aforementioned two limitations and the uptake of process mining in the healthcare domain have been the reason for writing this *SpringerBrief* about process mining in healthcare. To this end, we present a *healthcare reference model* which outlines all the different classes of data that are potentially available for process mining and the relationships between these classes. Given this reference model, it is possible to reason about application opportunities for process mining, e.g., we will discuss several kinds of analyses that can be performed. This enables us to answer the following question: *What are the potential applications of process mining within hospitals?*

When applying process mining in hospitals, typically several data quality issues need to be tackled. For example, problems may exist related to timestamps in event logs, imprecise activity names, and missing events. Therefore, we also elaborate on *data quality issues*. In total 27 quality issues that may hamper the analysis of care processes based on event data were identified. We also provide *guidelines* to overcome these problems.

In the remainder, we provide an extensive overview of the issues and opportunities related to applying process mining in the healthcare domain. As such, a basis is provided for governing and improving the processes within a hospital.

Figure 1.5 shows an overview of this *SpringerBrief*. Starting point is a HIS (or similar system) that is supporting healthcare professionals. The healthcare reference model describes over 120 classes of information stored in a typical HIS. The reference model facilitates the search for relevant data and the ETL (Extract, Transform and Load) process. The resulting event logs can be used by a wide range of process discovery algorithms. For example, process models may be discovered that show what actually happens thus providing valuable insights. Existing artifacts like guidelines can be combined with event data to diagnose deviations. Opportunities to further exploit event data are endless, e.g., detecting bottlenecks or predicting capacity problems.

The remainder of this *SpringerBrief* is organized as follows. Chapter 2 discusses what kinds of healthcare processes can be analyzed using process mining. Therefore, a classification of healthcare processes is presented which gives an overview of the kinds of processes that can be found within the healthcare domain. Next, in Chap. 3 an introduction to process mining is given. In Chap. 4, the *healthcare reference model* is introduced. In Chap. 5, based on the reference model, the possibilities of process mining within a typical hospital will be illustrated. Chapter 6 lists common data quality issues and provides guidelines for logging. Finally, in Chap. 7 a short

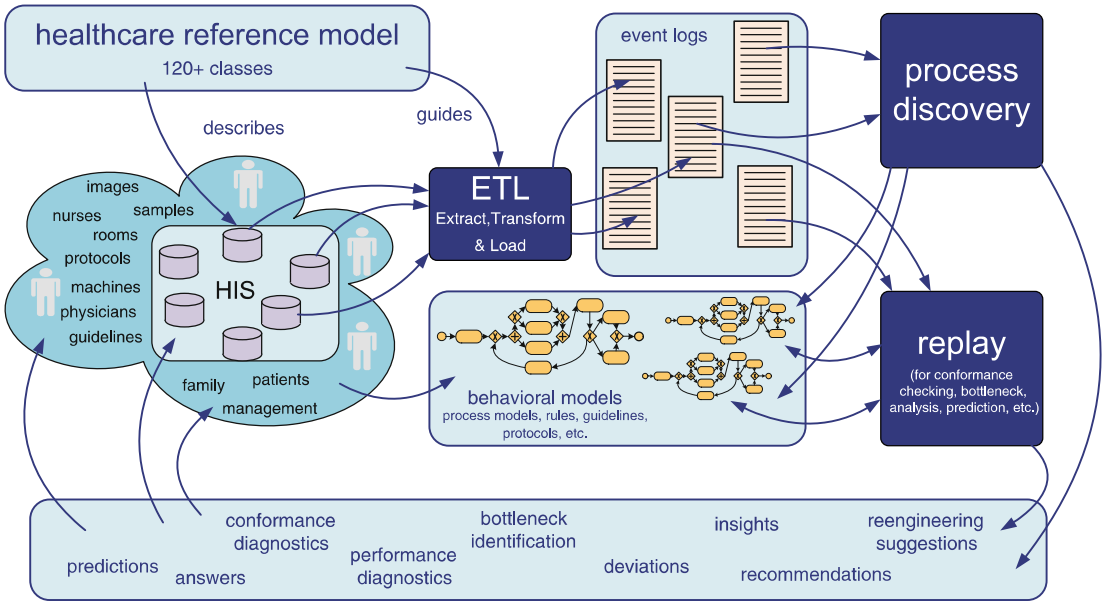

Fig. 1.5 Process mining in healthcare. Note that the healthcare reference model is used as a starting point for locating the data and extracting event logs

summary is given. Also, a vision for the application of process mining in healthcare based on the findings in this *SpringerBrief* is provided.

References

1. OECD. *Health at a Glance 2013: OECD Indicators*. OECD Publishing, 2013
2. L. Siciliani, M. Borowitz, and V. Moran. Waiting Time Policies in the Health Sector: What Works? Technical report, OECD Health Policy Studies, OECD Publishing, 2013
3. W.M.P. van der Aalst. Data Scientist: The Engineer of the Future. In K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, editors, *Proceedings of the I-ESA Conference*, volume 7 of *Enterprise Interoperability*, pages 13–28. Springer, 2014
4. C. Howard, D.C. Plummer, Y. Genovese, J. Mann, D.A. Willis, and D.M. Smith. The Nexus of Forces: Social, Mobile, Cloud and Information. http://www.gartner.com, 2012
5. T.H. Davenport and D.J. Patil. Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, pages 70–76, October 2012
6. W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011
7. W.M.P. van der Aalst, A.J.M.M. Weijters, and L Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004
8. W.M.P. van der Aalst, A. Adriansyah, and B.F. van Dongen. Replaying History on Process Models for Conformance Checking and Performance Analysis. *WIREs Data Mining and Knowledge Discovery*, 2(2):182–192, 2012
9. M. Song and W.M.P. van der Aalst. Towards Comprehensive Support for Organizational Mining. *Decision Support Systems*, 46(1):300–317, 2008

Chapter 2
Healthcare Processes

Abstract Process mining can be used to improve compliance and performance in hospitals and other care organizations. Before analyzing event data, we first provide an overview of the different types of care processes. We distinguish three levels of care: primary, secondary, and tertiary. We characterize five types of healthcare processes and link these to four basic types of data science questions: (a) What happened?, (b) Why did it happen?, (c) What will happen?, and (d) What is the best that can happen? Such questions can be answered using process mining. Using the characteristics of care processes, different questions may be posed. For example, the level of variability may influence the selection of the most suitable process mining technique.

Keywords Taxonomy of healthcare processes · Process mining · Healthcare · Data science · Characterizing healthcare processes

Healthcare can be seen as the diagnosis, treatment, and prevention of diseases in order to improve a person's wellbeing. Although healthcare is typically associated to hospitals, there are many care processes in other types of organizations. Various professionals may be involved in these processes. Examples of such professionals are general practitioners, dentists, midwives, and physiotherapists. Also, care is provided at home, rehabilitation centers, and nursing homes. Next to that, when looking at literature, typically it is indicated that healthcare processes are highly dynamic, complex, ad-hoc, and are increasingly multi-disciplinary [1]. Nevertheless, processes of different complexity and duration (up to several months) can be identified. This chapter provides a brief characterization of the spectrum of care processes encountered.

2.1 Different Levels of Care

Obviously, many different kinds of healthcare processes exist having different execution characteristics. Therefore, in this chapter, we focus on indicating which kinds of healthcare processes exist and the typical process characteristics that distinguish these processes. In the end, a classification is provided outlining the main kinds of healthcare processes.

R.S. Mans et al., *Process Mining in Healthcare*,
SpringerBriefs in Business Process Management, DOI 10.1007/978-3-319-16071-9_2

Regarding the organization of care, there is a distinction into three levels of care. Each level corresponds to particular patient needs [2]. First, *primary care* involves common health problems (e.g. sore throats and hypertension) and preventive measures (e.g. vaccinations or electrocardiography) that account for 80–90 % of visits to a physician or other caregiver. So, primary care refers to the work of healthcare professionals who act as a first point of consultation for all patients within the healthcare system [3]. As such, it is the basis for referrals to secondary and tertiary level care.

At the next level, within *secondary care*, problems are handled that require more specialized clinical expertise [2] (e.g. a patient with acute renal failure). In comparison to primary care, services are provided by physicians and other health professionals who generally do not have first contact with patients. Moreover, secondary care is usually short-term, involving sporadic consultation from a specialist to provide expert opinion and/or surgical or other advanced interventions that primary care physicians (PCPs) are not equipped to perform [4]. Secondary care thus includes hospitalization, routine surgery, specialty consultation, and rehabilitation [4]. Note that secondary care is not necessarily only provided within a hospital. Many professionals work outside hospitals such as physiotherapists or psychiatrists.

Finally, *tertiary care* involves the management of rare and complex disorders [2]. This care is usually provided for inpatients and on referral from a primary or secondary care medical professional. Examples of tertiary care are trauma care, burn treatment, neonatal intensive care, tissue transplants, and open heart surgery. Typically, tertiary care is institution-based, highly specialized, and technology-driven [2]. Much of this kind of care is provided in large teaching hospitals, especially university-affiliated teaching hospitals [2].

Within [5], *emergency care* is considered as another level of care. This kind of care is provided by emergency medicine professionals. Their mission is to evaluate, manage, treat, and prevent unexpected illness and injury. Emergency physicians provide rapid assessment and treatment of any patient with a medical emergency. In addition, they are responsible for the initial assessment and care of any medical condition that a patient believes requires urgent attention, and they provide medical care for individuals who lack access to other kinds of care. Commonly, three levels of care are distinguished within emergency care. First, there are *non-urgent* patients which typically require primary care. Second, there are *emergent* patients which have immediate life of limb threatening problems. Third, there are *urgent* patients which fall in between the two other levels of care [6]. Clearly, emergent patients require immediate treatment whereas for non-urgent this is not necessarily needed.

2.2 Classification of Healthcare Processes

From the discussion above it becomes clear that patients' care processes may differ substantially. For patients in the same homogeneous patient group the process execution may be comparable but in case of complex patients, the accompanying process

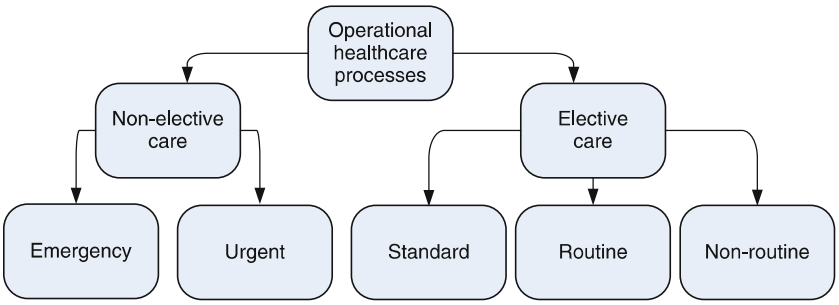

Fig. 2.1 Characterizing healthcare processes by outlining the main kinds of organizational healthcare processes

may exhibit many different execution outcomes. In order to come to a good understanding of the different characteristics of healthcare processes, Fig. 2.1 proposes a classification of the main kinds of healthcare processes that may coexist. Moreover, this classification is used to indicate on which kinds of processes the focus will be on in this book.

First of all, we only refer to care which is directly related to or provided for patients. In particular, the focus is on operational processes. These processes are concerned with the *logistics* of work processes. This involves the medical steps that need to be done together with the necessary preparations for these steps (e.g. the making of an appointment and the reservation of a room). As such we do not look into the process of individual decisions that are made by medical professionals with regard to diagnosis and treatments of patients. Also, for the activities within a process we do not look into the medical interpretation of them. Although these activities are included in a process mining analysis as individual events, we do not consider the results of these activities (e.g. the outcome of a blood test or X-ray) or any interpretation associated. In other words: *we focus on the orchestration and management of the care processes* rather than individual activities.

Given the above mentioned focus, in Fig. 2.1, operational processes are further subdivided into two main classes. *Elective care* relates to care for which it is medically sound to postpone treatment for days or weeks [7]. Conversely, *non-elective care* represents patient for whom medical treatment is unexpected and needs to be planned on short notice [7].

Elective care can range from processes that are completely standardized till processes for which a huge amount of variation exists. According to Lillrank et al. [8], a further division can be made into three subclasses. First, for *standard* processes a standardized treatment path exists which defines the different activities in the process and their timing. Given targets should be achieved if a treatment path is meticulously followed. Second, for *routine* processes, the overall outcome of the process is usually known. However, different process paths may be followed during treatment. Finally, within *non-routine* processes, the physician proceeds in a step-by-step way, checking the patient's reaction to an individual treatment and deciding about the steps to be taken next [9, 10]. Often, the decision about the next steps to be taken is not made by just one physician. When complex care needs to be delivered, cooperation between

various physicians across different medical specialties and departments is needed to decide on and execute parts of the individual patient's care plan.

Finally, non-elective care can be further subdivided into two classes being emergency care and urgent care. *Emergency care* has to be performed immediately. In contrast, *urgent care* can be postponed for a short time (i.e. multiple days). The non-urgent patients mentioned earlier for emergency care [5] are considered to be elective care rather than non-elective care.

Considering the classification presented in Fig. 2.1 it is clear that many challenges exist related to the application of process mining on healthcare processes. For some processes, the discovered model may be relatively simple (standard care, routine care, and urgent care) whereas for other processes, the discovered process is much more spaghetti-like (emergency care and non-routine care). Furthermore, although processes may be standardized and homogeneous patient groups exist, still some variation may exist within a process. This relates to inter-physician variation and inter-practice variation within hospitals. With process mining such variations become visible.

2.3 Four Types of Questions

As mentioned in the introduction we distinguish four types of data science questions. After characterizing the different types of healthcare processes, we provide some example questions for each of the four categories.

- *What happened?*
  - What is the typical treatment of patients having acute myeloid leukemia?
  - How and when are patients transferred to an academic hospital?
  - What is the typical working day of a surgeon?
  - How is the new telehealth system used?
- *Why did it happen?*
  - What caused the unusual amount of incidents in the department?
  - Why was the service level agreement not reached?
  - Why did people stop using the telehealth system?
  - What caused the long waiting list?
- *What will happen?*
  - When will this patient be dismissed?
  - Is this patient likely to deviate from the normal treatment plan?
  - How many beds are needed tomorrow?
  - Is it possible to handle these five new cases in time?

- *What is the best that can happen?*
  - Which check should be done first to reduce flow time?
  - How many physicians are needed to reduce the waiting list by 50 %?
  - Does it help to do these tests concurrently?
  - How to redistribute the workload over the three surgeons?

As we will see in the remainder, such questions can be answered using process mining. In the next chapter an introduction to process mining is given. Chapter 4, introduces the healthcare reference model. Based on this reference model, Chap. 5 illustrates the array of analysis possibilities provided by process mining.

References

1. A. Rebuge and D.R. Ferreira. Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining. *Information Systems*, 37(2), 2012
2. K. Grumbach and T. Bodenheimer. The Organization of Health Care. *The Journal of the American Medical Association*, 273(2):160–167, 1995
3. "World Health Organization". A Glossary of Terms for Community Health Care and Services for Older Persons. WHO Centre for Health Development: Ageing and Health Technical Report 5, "WHO", 2004
4. L. Shi. The Impact of Primary Care: A Focused Review. *Scientifica*, 2012:22, 2012
5. B. Starfield. Is Primary Care Essential? *The Lancet*, 344:1129–1133, 1994
6. S.M. Schneider, G.C. Hamilton, P. Moyer, and J.S. Stapczynski. Definition of Emergency Medicine. *Academic Emergency Medicine*, 5(4):348–351, 1998
7. D. Gupta and B. Denton. Appointment Scheduling in Health Care: Challenges and Opportunities. *IIE Transactions*, 40(9):800–819, 2008
8. P. Lillrank and M. Liukko. Standard, Routine and Non-Routine Processes in Health Care. *International Journal of Health Care Quality Assurance*, 17(1):39–46, 2004
9. G de Vries, J W M Bertrand, and J M H Vissers. Design Requirements for Health Care Production Control Systems. *Production Planning & Control*, 10(6):559–569, 1999
10. J. Vissers and R. Beech. Chain Logistics: Analysis of Care Chains. In J. Vissers and R. Beech, editors, *Health Operations Management: Patient Flow Logistics in Health Care*, Routledge Health Management Series, pages 70–83. Routledge, 2005

Chapter 3
Process Mining

Abstract Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand-made or discovered automatically). This technology has become available only recently, but is mature enough to be applied to care processes of any type and of any complexity. The process-mining spectrum is broad and includes techniques for process discovery, conformance checking, prediction, and bottleneck analysis. Traditional data-mining approaches are not process-centric. Input for data mining is typically a set of records and the output is a decision tree, a collection of clusters, or frequent patterns. Process mining starts from events and the output is related to an end-to-end process model. Data mining tools can be used to support particular decisions in a larger process. However, they cannot be used for process discovery, conformance checking, and other forms of process analysis. Therefore, process mining is needed to improve compliance and performance in hospitals in a systematic manner.

Keywords Process discovery · Conformance checking · Process mining · Event logs · Healthcare · Business process management

Given an event log, the goal of process mining is to extract process knowledge (e.g. process models) in order to discover, monitor, and improve real processes [1]. These event logs may originate from a wide range of systems. Examples of these systems are Business Process Management (BPM) systems (e.g. *BPM|One*, *Filenet*), ERP systems (e.g. *Microsoft Dynamics NAV* and *SAP*), Product Data Management (PDM) systems (e.g. *Windchill*), and Hospital Information Systems (e.g. *i.s.h.med*, *iSOFT*, *ChipSoft*, *McKesson*, and *Epic*). In later chapters we will focus on the typical event data found in the latter class of systems. However, first we provide a brief overview of process mining.

R.S. Mans et al., *Process Mining in Healthcare*,
SpringerBriefs in Business Process Management, DOI 10.1007/978-3-319-16071-9_3

3.1 Event Data and Process Models

Within an event log certain information needs to be present in order to apply process mining. Figure 3.1 depicts the typical information that needs to be available in terms of a UML class diagram [2]. Furthermore, the relation between an event log and a process model is visualized. Three levels are identified: model level, instance level, and event level. The model level and the event level typically *exist independent of one another*; people make process models without relating to (raw) data in the information system and processes generate data while being unaware of the process models that may exist.

The instance level consists of *cases* and *activity instances*. These connect *processes* and *activities* in the model to *events* in the log. There are several important relationships between these five concepts together with their cardinalities. At the model level, a process may have an arbitrary number of activities, but each activity belongs to precisely one process (cardinality `1..*`). An event log contains information about a single process. Such a process may consist of multiple cases but each every case only belongs to precisely one process (cardinality `1..*`). Furthermore, each event in the log belongs to a single case (cardinality `1..*`). Events are related to activities. In particular, every event corresponds to one activity instance (cardinality `1..*`) whereas each activity instance refers to precisely one activity (cardinality `1..*`).

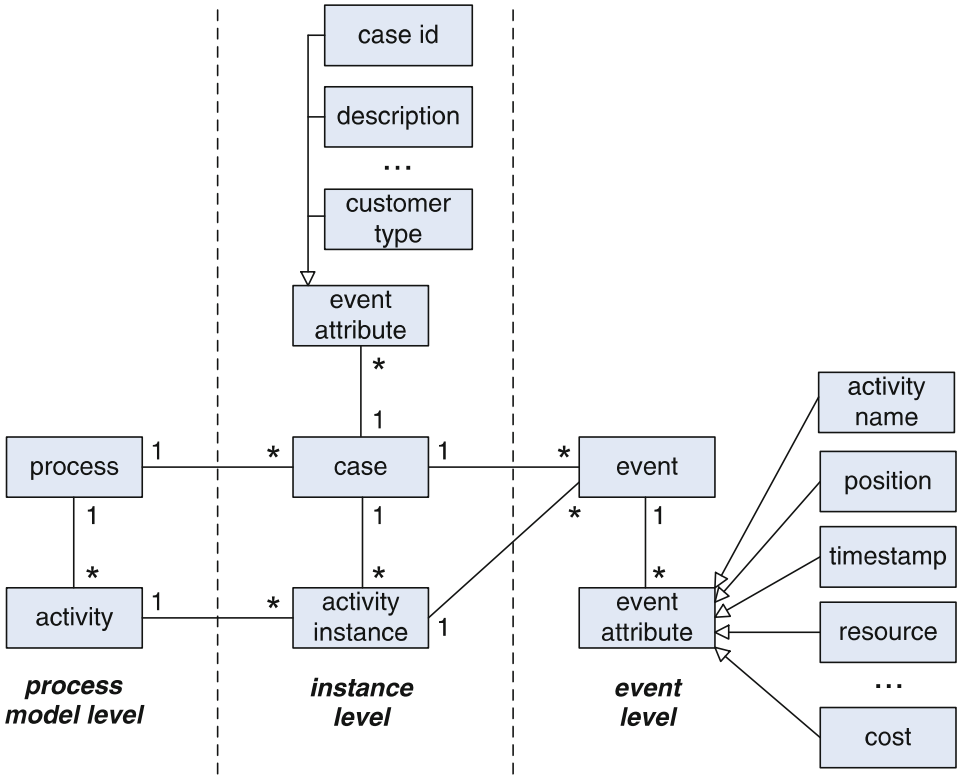

Fig. 3.1 The typical information that needs to be present in an event log. Additionally, the relation between a process model and a event log is depicted

Note that multiple activity instances may exist for one activity and that for the same activity instance there may be multiple events. Furthermore, every activity instance belongs to precisely one case (cardinality `1..*`).

For cases and events additional information may exist. This information is contained in attributes (cardinality `1..*`). Each attribute consists of a name and a value (e.g. "(resource, Ferdinand)"). For an event, two attributes deserve some special attention. In order to discover causal dependencies in process models, events need to be ordered. This requirement is satisfied if all events have a *timestamp* representing the time at which the event occurred. In case this information is not available, a *position* attribute is needed which specifies the index of an event in a case. Note that timestamp information can be used for calculating performance properties of the process. The "resource" attribute (performer of the event) and the "cost" attribute (cost of the activity) can be used for inferring additional process knowledge.

An example log is shown in Fig. 3.2a. The table contains 16 events for 3 cases. For example, for the case with id "1", subsequently the activities "Fist Visit", "Surgery", "Second Visit", "Radiotherapy", "Chemotherapy" and "Evaluate" have been performed. Here, the "Fist Visit" event has id "4798669", is performed by "Pete" at "02/06/2014 14:00:00", and has cost "150".

Obliviously, process mining is related to data mining, machine learning and Business Intelligence (BI). These also aim at knowledge discovery, performance measurement, and prediction [3]. However, process mining is process-centric, an aspect largely ignored by mainstream data mining, machine learning and BI techniques. Process mining starts with making unknown (or only partially known) processes explicit in terms of end-to-end process models (not just patterns).

Process mining is also closely related to *Business Process Management* (BPM). Figure 3.3 shows a variant of the classical BPM lifecycle. In the *(re)design phase*, a process model is designed. This model is transformed into a running system in the *implementation/configuration phase*. If the model is already in executable form and a WFM or BPM system is already running, this phase may be very short. However, if the model is informal and needs to be hardcoded in conventional software, this phase may take substantial time. After the system supports the designed processes, the *run & adjust phase* starts. In this phase, the processes are enacted and adjusted when needed. In the run & adjust phase, the process is not redesigned and no new software is created; only predefined controls are used to adapt or reconfigure the process.

Next to the BPM lifecycle, Fig. 3.3 also shows the two main types of analysis: *model-based analysis* and *data-based analysis*. While the system is running, event data are collected. These data can be used to analyze running processes, e.g., discover bottlenecks, waste, and deviations. This is input for the redesign phase. During this phase process models can be used for analysis. For example, simulation is used for what-if analysis or the correctness of a new design is verified using model checking. In recent years the focus in BPM shifted from purely model-based analysis to data-based analysis, thus explaining the growing interest in process mining.

(a)

Case id	Event id	Properties			
		Timestamp	Activity	Resource	Cost
1	4798669	02/06/2014 14:00	First Visit	Pete	150
	4798670	07/06/2014 11:00	Surgery	Rose	55
	4798677	09/06/2014 17:00	Second Visit	Pete	150
	4798679	12/06/2014 12:15	Radiotherapy	Alfred	200
	4798680	14/06/2014 12:15	Chemotherapy	Michael	300
	4798685	17/06/2014 10:00	Evaluate	Pete	175
2	7777171	05/06/2014 09:15	First Visit	John	120
	7777195	13/06/2014 14:30	Surgery	Nick	610
	7777189	15/06/2014 15:45	Second Visit	John	120
	7777179	27/06/2014 11:45	Chemotherapy	Michael	300
	7777195	29/06/2014 14:30	Radiotherapy	Alfred	200
	7777185	08/07/2014 08:15	Evaluate	Pete	180
3	7932191	12/06/2014 14:45	First Visit	John	120
	7932192	23/06/2014 11:00	Surgery	Rose	55
	7932196	25/06/2014 08:00	Second Visit	John	120
	7932192	01/07/2014 16:00	Radiotherapy	Alfred	200
	7932193	07/07/2014 17:30	Chemotherapy	Michael	300
	7932196	21/07/2014 08:00	Evaluate	John	185

(b)

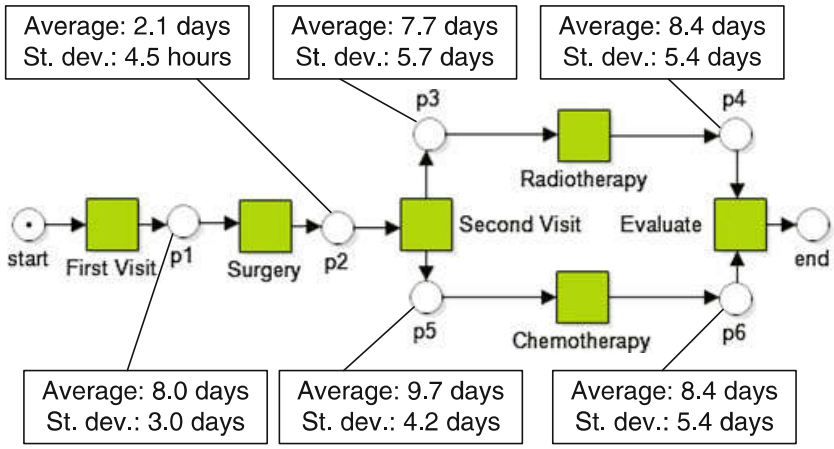

Fig. 3.2 Example of an event log together with discovered process knowledge. **a** A fragment of an event log: each line corresponds to an event. **b** Discovered Petri net for the event log fragment. Additionally, for the places some statistics are shown concerning the time spent in the places

Figure 3.4 again shows that process mining is the missing link between analysis techniques that focus on process models without considering the actual event data and classical data-oriented analysis with no attention to end-to-end processes. The figure also shows that process mining can be used to answer both *compliance-related* (e.g., where and why do physicians deviate?) and *performance-related* (e.g., where are the main bottlenecks and how to remove them?) *questions*.

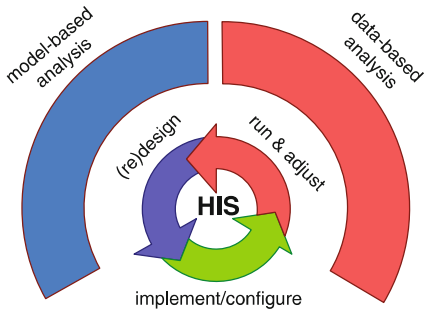

Fig. 3.3 The BPM lifecyle and the role of data-based analysis versus model-based analysis

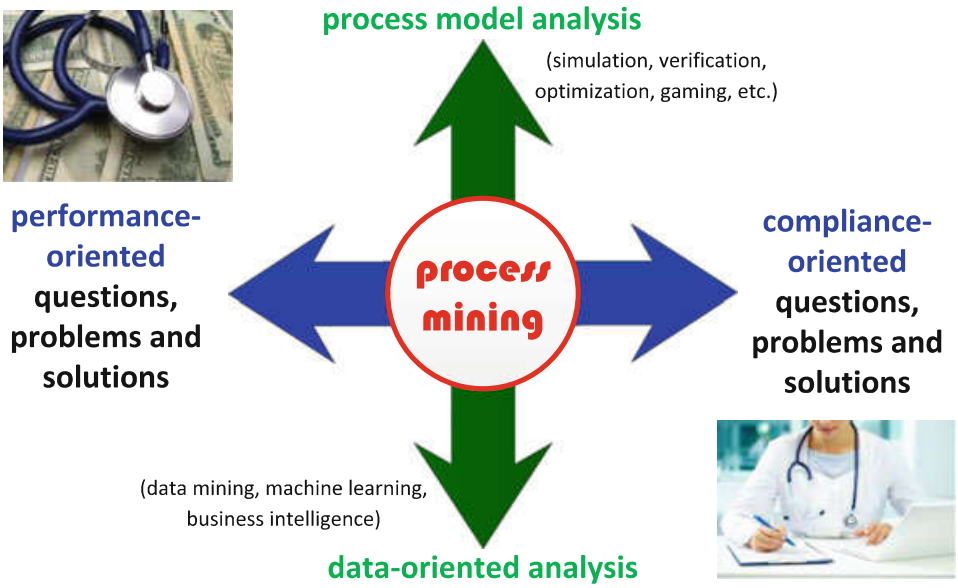

Fig. 3.4 Process mining aims to answer performance-related and compliance-related questions

3.2 Three Types of Process Mining

In general, three main types of process mining can be distinguished (see Fig. 3.5). **Discovery**: Here the focus is on inferring process models (e.g. a Petri net or a BPMN model) that are able to describe the observed behavior. For example, the inferred model may describe the typical steps that are taken within a process.

Figure 3.2b shows a Petri net that is discovered for the event log of Fig. 3.2a. As can be seen for all cases, the process starts with a "First Visit". Next, a "Surgery" and the "Second visit" is performed. After that both "Radiotherapy" and "Chemotherapy" are performed. Note that these two activities may be performed in any order. Finally, an evaluation (task "Evaluate") is taking place. Note that also models describing the organizational or data perspectives may be discovered.

Conformance: for a given model it is checked whether it conforms to observed behavior in the event log. In case there are deviations between the model and the event log these are identified such that they can be further analyzed (e.g. activities in the model that cannot be found in the event log or the other way around).

For example, for the traces shown in Fig. 3.2a it can be seen that all their events can be successfully replayed in the model of Fig. 3.2b. However, consider the trace with events "First visit", "Surgery", "Second visit", "Radiotherapy" and "Evaluate". When replaying this trace it is found that the "Chemotherapy" event is missing. Next, for the trace with events "First visit", "Surgery", "Second Visit", "Chemotherapy", "Lab test", "Radiotherapy", and "Evaluate" it can be seen that an lab test is done in between the "Second visit" and "Evaluate" activities.

Enhancement: information extracted from the log is projected onto the model. Note that here it is assumed that already a model exists (either discovered or made by hand). Using the seminal notion of alignments, trace in the log can be connected to paths in the model (even in case of deviations). Alignments can be used to "repair" a process model, i.e., the model is enhanced by making it closer to reality, but still retaining as much as possible from the original model. Alignments can also be used to enrich the model with additional perspectives (times, costs, risks, decisions, resource usage, etc.). For example, in the example event log also timestamp information is found for the events. This can be used for calculating performance information concerning the discovered process shown in Fig. 3.2. That is, for each place in the model the average time that is spent by a token in that place is indicated. Additionally, the standard deviation is given. For example, the average time in between a "Surgery" and the "Second Visit" is 2.1 days (standard deviation: 4.5 h).

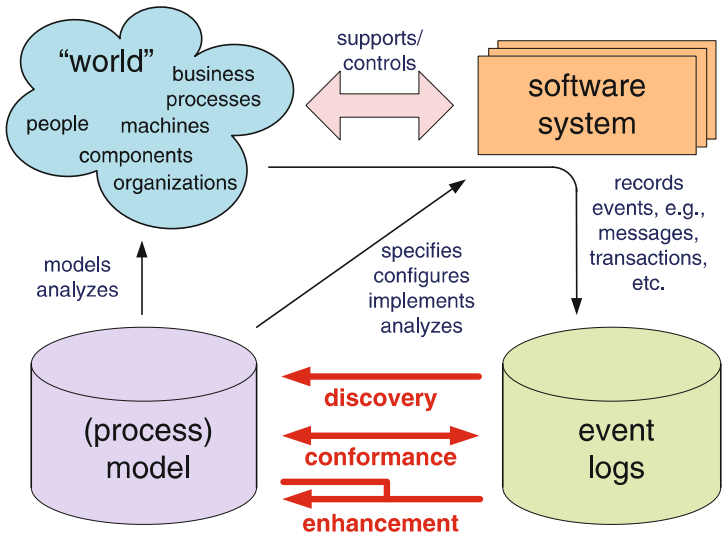

Fig. 3.5 Three types of process mining: (1) Discovery, (2) Conformance, and (3) Extension

3.3 The Process Mining Spectrum

Thus far we identified three main types of process mining: *discovery*, *conformance*, and *enhancement*. However, this does not reflect the broadness of the process mining spectrum. Orthogonal to there three types of process mining are perspectives such as the *control-flow perspective* ("How?"), the *organizational perspective* ("Who?"), and the *case/data perspective* ("What?"). Moreover, analysis can be done *online* or *off-line*.

Figure 3.6 (taken from [1]) shows the so-called refined process mining framework. Data in event logs are partitioned into "*pre mortem*" and "*post mortem*" *event data*. "Post mortem" event data refer to information about cases that have completed, i.e., these data can be used for process improvement and auditing, but not for influencing the cases they refer to. "Pre mortem" event data refer to cases that have not yet completed. If a case is still running, i.e., the case is still "alive" (pre mortem), then it may be possible that information in the event log about this case (i.e., current data)

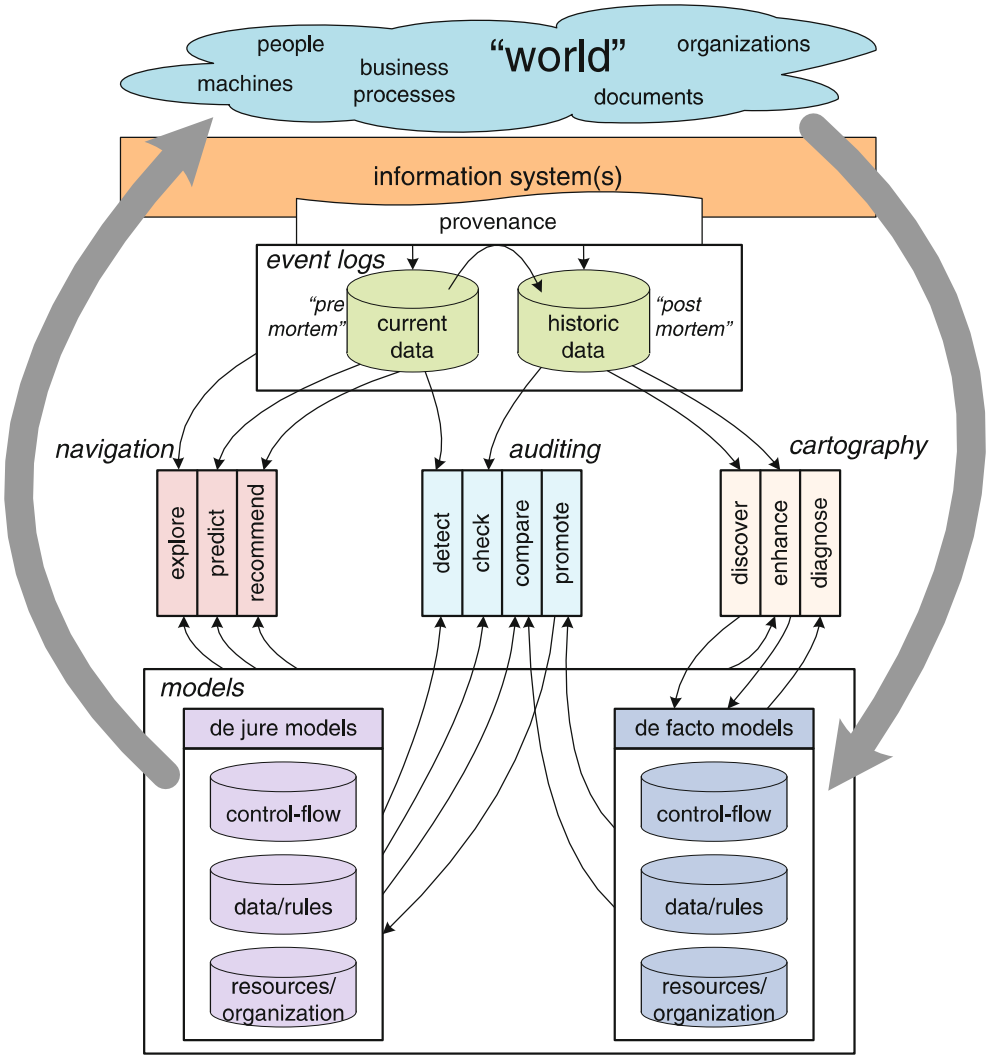

Fig. 3.6 Process mining framework providing an overview of the process mining spectrum [1]

can be exploited to ensure the correct or efficient handling of this case. The refined process mining framework also distinguishes between two types of models: "*de jure models*" and "*de facto models*". *A de jure model is normative*, i.e., it specifies how things should be done or handled. For example, a process model used to configure a BPM system is normative and forces people to work in a particular way. *A de facto model is descriptive* and its goal is not to steer or control reality. Instead, de facto models aim to capture reality adequately.

By combining the different types of event data ("pre mortem" or "post mortem"), the different types of perspectives (control-flow, organizational, data, costs, etc.), and the different types of models ("de jure" and "de facto models") the broadness of the field becomes obvious. Figure 3.6 lists ten process-mining related activities.

More and more organizations are adopting process mining as a means to improve operational performance and conformance. For example, the web site of the IEEE Task Force on Process Mining[1] which lists over 15 successful case studies in industry. Process mining has also been applied in several Dutch hospitals (Isala Hospital Zwolle, Maastricht University Medical Center, Academic Medical Center Amsterdam, Catharina Hospital Eindhoven, Mental Healthcare Institute Eindhoven, Albert Schweitzer Hospital Dordrecht, etc.).

3.4 Tool Support

The *ProM* framework[2] is the de facto standard for process mining aiming to cover the whole spectrum shown in Fig. 3.6 [1, 4]. ProM is a "plug-able" environment for process mining using MXML, SA-MXML, or XES as input format. Import tools like ProMimport and XESame can be used to convert data from various sources into event logs. Moreover, tools like Disco and ProM can also read CSV files and interpret these as event logs. ProM provides hundreds of plug-ins supporting the ten process mining related activities shown in Fig. 3.6.

The uptake of process mining is not only illustrated by the growing number of papers and plug-ins of the open source tool *ProM*, there are also a growing number of commercial analysis tools providing process mining capabilities, cf. *Disco* (Fluxicon), *Perceptive Process Mining* (Perceptive Software, before Futura Reflect and BPMone by Pallas Athena), *ARIS Process Performance Manager* (Software AG), *Celonis Process Mining* (Celonis GmbH), *ProcessAnalyzer* (QPR), *Interstage Process Discovery* (Fujitsu), *Discovery Analyst* (StereoLOGIC), and *XMAnalyzer* (XMPro). Many of the ideas developed in the context of ProM have been embedded in these commercial tools.

Mainstream BI software like IBM Cognos Business Intelligence (IBM), Oracle Business Intelligence (Oracle), SAP BusinessObjects (SAP), WebFOCUS (Information Builders), MS SQL Server (Microsoft), MicroStrategy (MicroStrategy),

[1] www.win.tue.nl/ieeetfpm/doku.php?id=shared:process_mining_case_studies.

[2] www.processmining.org.

NovaView (Panorama Software), QlikView (QlikTech), SAS Enterprise Business Intelligence (SAS), TIBCO Spotfire Analytics (TIBCO), Jaspersoft (Jaspersoft), and Pentaho BI Suite (Pentaho) are *not* process oriented and therefore less suitable for answering process mining questions. Similar comments can be made about data mining tools. Although both process mining and data mining start from data, data mining techniques are typically not process-centric and do not focus on event data. For data mining techniques the rows (instances) and columns (variables) can mean anything. For process mining techniques, we assume event data where events refer to process instances and activities. Moreover, the events are ordered and we are interested in end-to-end processes rather than local patterns. End-to-end process models and concurrency are essential for process mining. Moreover, topics such as process discovery, conformance checking, and bottleneck analysis are not addressed by traditional data mining techniques and tools (Fig. 3.7).

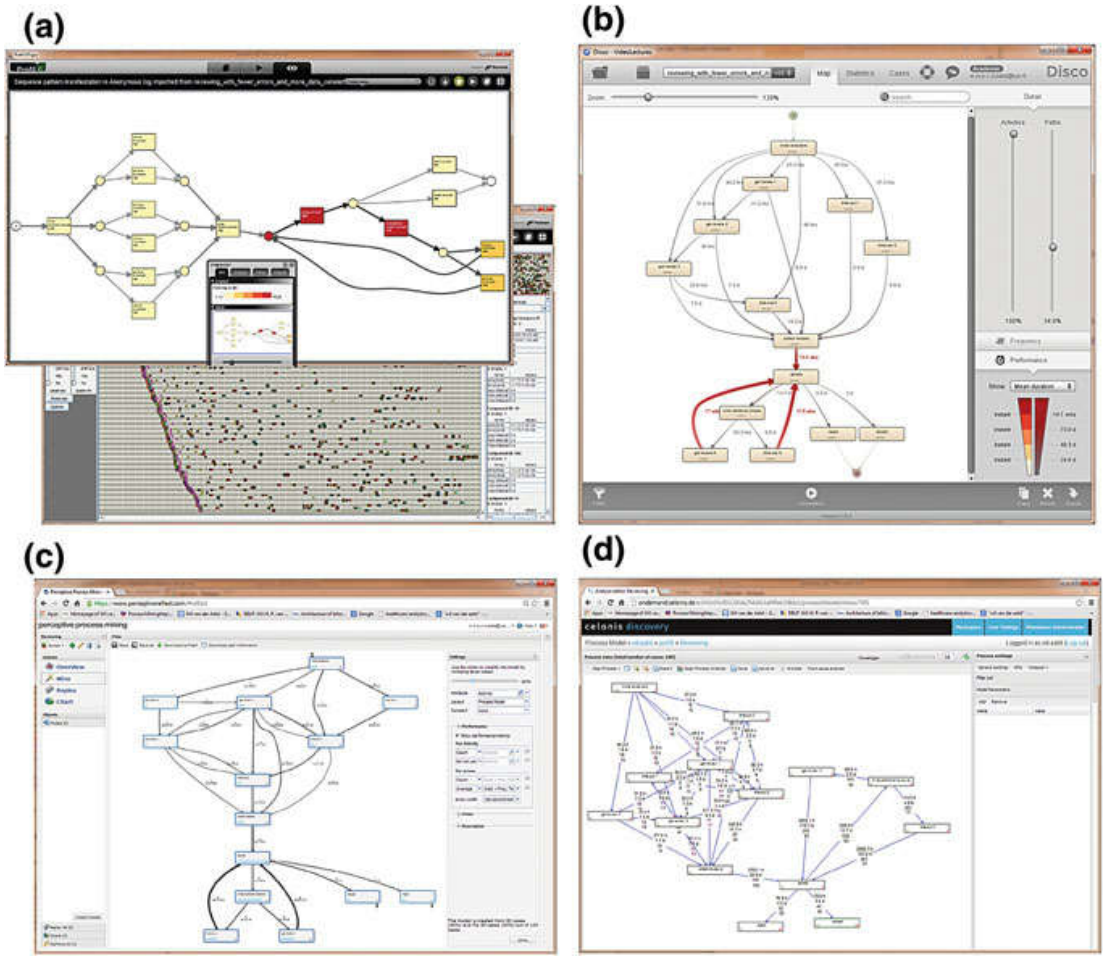

Fig. 3.7 Four screenshots of different tools analyzing the same event log: **a** *ProM* (open source), **b** *Disco* (Fluxicon), **c** *Perceptive Process Mining* (Perceptive Software), and **d** *Celonis Process Mining* (Celonis GmbH)