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

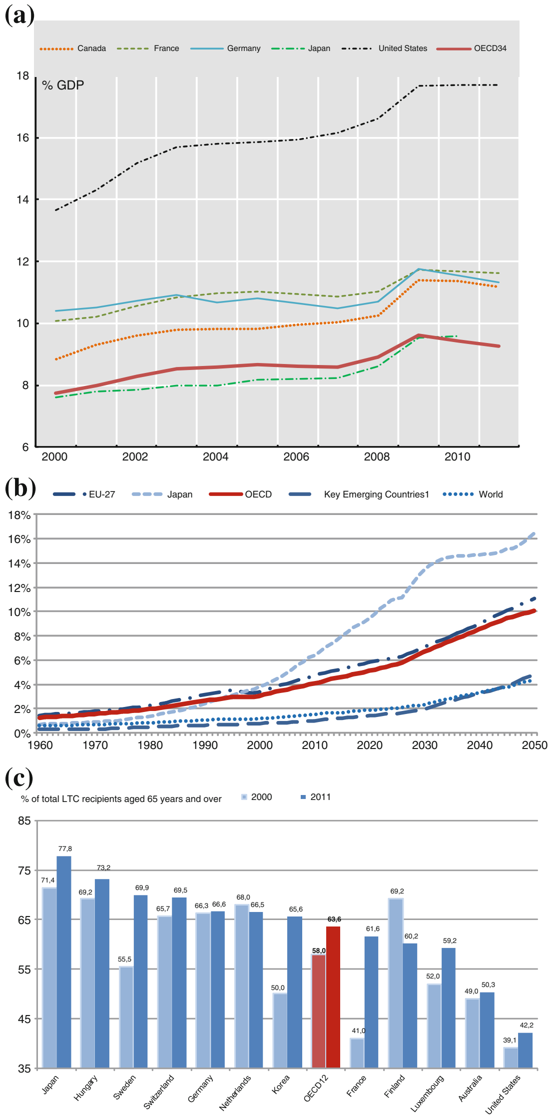

Fig. 1.1 Within healthcare, costs are rising, people are aging, and the demand for care is increasing. **a** Total health expenditure as a share of GDP, 2000–2011. *Source* OECD Health Statistics 2013, http://dx.doi.org/10.1787/health-data-en. **b** Trends in the share of the population aged over 80 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

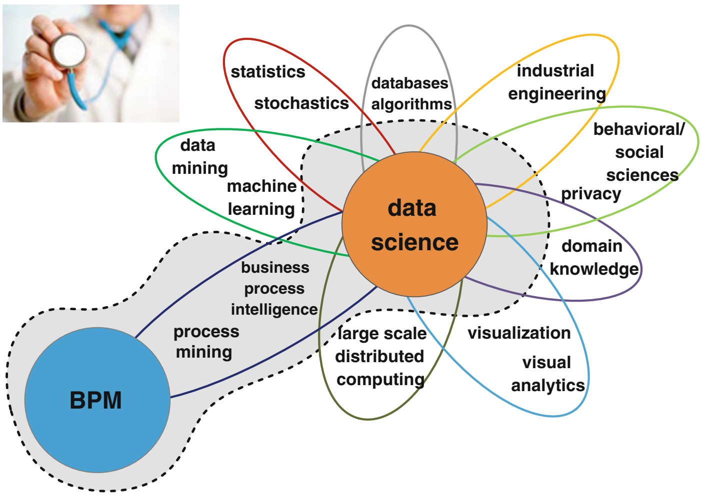

Fig. 1.2 Data science skills that should be combined to realize more efficient and effective care processes

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.

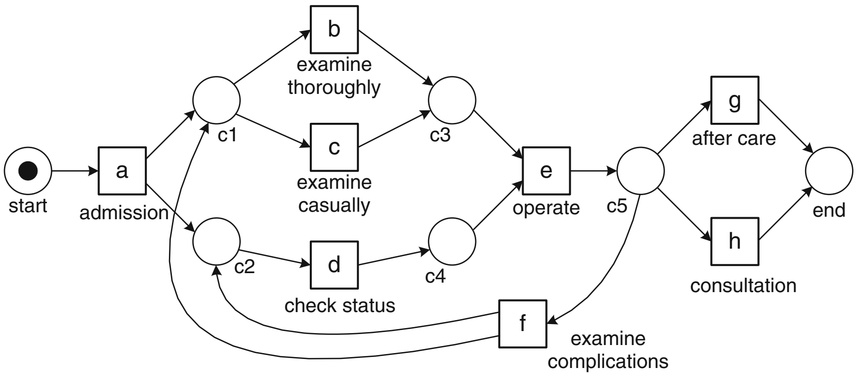

Fig. 1.3 Process model only showing the control-flow. The model is not intended to be realistic and only aims to show the different control-flow constructs in a healthcare setting

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.

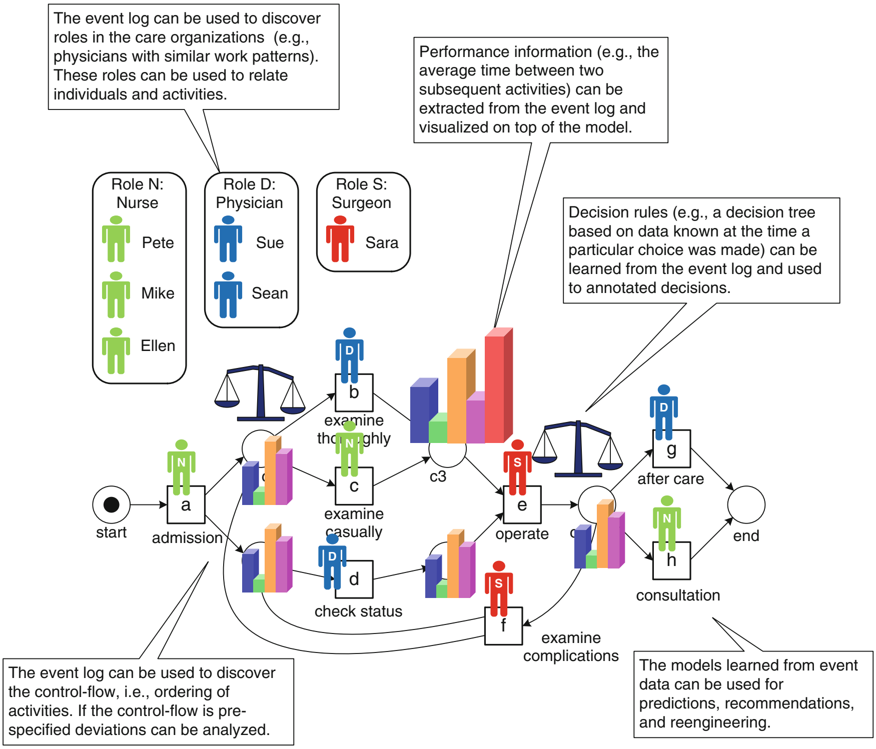

Fig. 1.4 Process mining is not only used to learn the process as it is actually executed: It is also used to understand deviations, to analyze bottlenecks, and to monitor organizational behavior

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

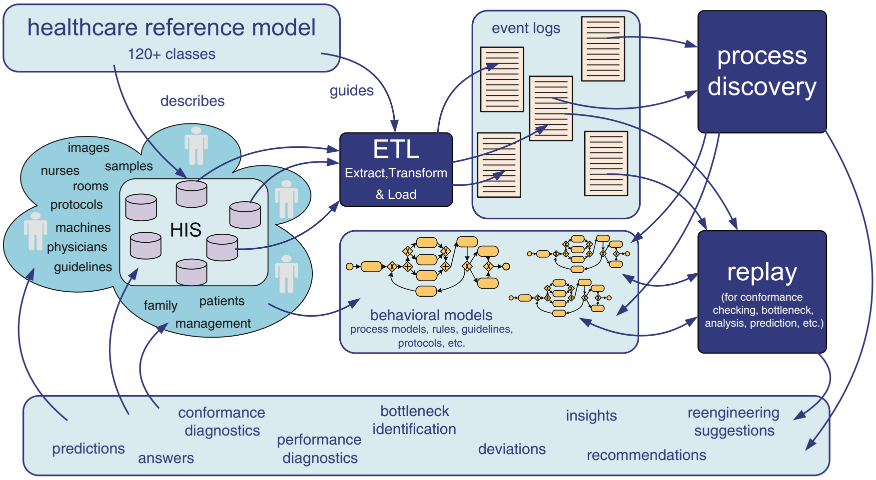

Fig. 1.5 Process mining in healthcare. Note that the healthcare reference model is used as a starting point for locating the data and extracting event logs

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

Springer

http://www.springer.com/978-3-319-16070-2

Process Mining in Healthcare
Evaluating and Exploiting Operational Healthcare Processes
Mans, R.S.; van der Aalst, W.; Vanwersch, R.J.B.
2015, X, 91 p. 43 illus., 6 illus. in color., Softcover
ISBN: 978-3-319-16070-2