Using indoor location system data to enhance the quality of healthcare event logs: opportunities and challenges

Niels Martin[0000−0003−3279−3853]

Hasselt University, Agoralaan Building D, 3590 Diepenbeek, Belgium
niels.martin@uhasselt.be

Abstract. Hospitals are becoming more and more aware of the need to manage their business processes. In this respect, process mining is increasingly used to gain insight in healthcare processes, requiring the analysis of event logs originating from the hospital information system. Process mining research mainly focuses on the development of new techniques or the application of existing methods, but the quality of all analyses ultimately depends on the quality of the event log. However, limited research has been done on the improvement of data quality in the process mining field, which is the topic of this paper. In particular, this paper discusses, from a conceptual angle, the opportunities that indoor location system data provides to tackle event log data quality issues. Moreover, the paper reflects upon the associated challenges. In this way, it provides the conceptualization for a new area of research, focusing on the systematic integration of an event log with indoor location system data.

Keywords: Event log · Data quality · Hospital information systems · Indoor location systems data · Process mining.

1   Introduction

The healthcare sector in general, and hospitals in particular, are confronted with challenges such as tightening budgets contrasted to increased care needs due to the aging population [4, 15, 19]. To face these challenges, hospitals are becoming increasingly aware of the need to manage their processes in order to improve them [15]. In this respect, process mining is gaining more attention as a way to gain insights in healthcare processes. Process mining is the extraction of knowledge from an event log containing process execution information from a process-aware information system such as a hospital information system (HIS).

Process mining research mainly focuses on the development of new techniques to extract knowledge from an event log or the innovative application of existing techniques [5]. However, consistent with the "garbage in - garbage out" principle, the quality of all process mining analyses ultimately depends on the quality of the event log used as an input [19]. Bose et al. [5] state that most real-life event logs struggle with issues such as incompleteness, noisiness, and imprecision. This

also holds in healthcare, where it is not always possible to extract high-quality data from a HIS [8]. Despite such observations, which are broadly shared given their inclusion in the Process Mining Manifesto [2], limited research has been done on the improvement of data quality within the process mining field. One research direction, which did not yet receive explicit attention, is considering the enrichment and improvement of event logs using other process-related data sources such as location data.

This paper discusses, from a conceptual angle, how indoor location system (ILS) data can be used to alleviate data quality issues present in an event log originating from a HIS. In healthcare, ILS systems are increasingly used for e.g. patient flow and staff workflow management [4]. Data generated by such systems provides information on the location of process participants such as patients, staff members, and medical equipment at a particular moment. This paper outlines which opportunities ILS data provides to tackle event log quality issues, but also reflects upon the associated challenges. In this way, it provides the conceptualization for a new research area, focusing on a systematic integration of an event log with ILS data. The resulting enhanced event log will contribute towards exploiting the full potential of process mining in healthcare practice.

This paper is structured as follows. Section 2 introduces the notions of an event log and ILS data. In Section 3, an overview of related work is provided. Section 4 details the potential of ILS data to alleviate event log quality issues. Despite these opportunities, several challenges are still ahead, as discussed in Section 5. The paper ends with a conclusion in Section 6.

2   Preliminaries

This section outlines the notions of an event log (Section 2.1) and ILS data (Section 2.2), which are the key data sources considered in this paper.

2.1   Event log

An event log is a data file containing process execution information. It consists of a collection of events, e.g. the completion of patient registration at the reception desk, associated to a case such as a patient. It minimally consists of an ordered set of events for each case, but typically also includes information such as a timestamp, and the resource associated to the event [1].

Table 1 illustrates the structure of an event log, where each line represents an event. For instance: the first line indicates that the registration of patient 103 by resource Mike started on April, 25th at 10:59:41. He completed this registration at 11:04:04, as shown in the second row of Table 1.

2.2   Indoor location system data

ILS data originates from an indoor location system (ILS), also referred to as a real-time location system. From a technical perspective, wireless Radio Frequency Identification (RFID) technology is used. Locations are determined using

Table 1: Illustration of event log structure

case id	timestamp	activity	transaction type	resource	...
...	...	...	...	...	...
103	25/04/2018 10:59:41	Register patient	start	Mike	...
103	25/04/2018 11:04:04	Register patient	complete	Mike	...
104	25/04/2018 11:04:04	Register patient	start	Mike	...
109	25/04/2018 11:06:22	Clinical examination	start	Judy	...
...	...	...	...	...	...

RFID tags, which can e.g. be integrated in patient identification bracelets or staff cards, and antennas which are deployed in a particular department [4]. For more technical details on an ILS, the reader is referred to [4].

Raw ILS data records the location of a process participant at a particular point in time. Technology provider's software often summarizes this raw data such that ILS data expresses time periods during which a process participant resided at a particular location [8, 23]. When this is not the case, preprocessing is required to obtain a dataset in the format exemplified in Table 2. For instance: the first line shows that the triage nurse with RFID-tag 1044 was present in the second triage room on April 25th from 12:58:06 until 13:22:14.

Table 2: Illustration of ILS data structure

tag id	type	location	start	end
...	...	...	...	...
1044	triage nurse	Triage room 2	25/04/2018 12:58:06	25/04/2018 13:22:14
7862	patient	Triage room 2	25/04/2018 13:12:14	25/04/2018 13:16:44
7809	patient	Box 12	25/04/2018 13:12:24	25/04/2018 13:52:03
1003	nurse	Box 12	25/04/2018 13:18:22	25/04/2018 13:26:02
...	...	...	...	...

3   Related work

Given the potential of process mining to gain profound insights in processes, it is increasingly studied in a healthcare context. Process mining methods are, amongst others, used to retrieve the activity order in healthcare processes [6, 9], to mine social networks [6, 18], and to check the conformance between an event log and a process model [15, 22]. A recent literature review on process mining in healthcare can be found in [21].

As for other data-oriented research domains, data quality should be one of the process mining community's prime concerns. Data quality is a multi-dimensional concept [24] which is studied in fields such as statistics, management, and computer science [3]. Literature provides several general frameworks to classify data quality issues [3, 14]. Moreover, dedicated data quality research has been done for specific types of data. For instance, Gschwandtner et al. [11] focus on time-related data and distinguish between quality issues present in a single dataset

and problems stemming from the combination of several datasets. Identified data quality issues include the start of a time interval being later than its end, and different timestamp structures in multiple datasets [11].

Even though insights from general data quality works are conceptually relevant for event logs, the particularities of process-related data warrant dedicated research efforts [24]. In this respect, the Process Mining Manifesto [2] defines five event log maturity levels, with increasing maturity levels implying improved process mining potential. While the maturity levels are rather generic, Bose et al. [5] identify 27 specific event log quality issues, which are grouped in four categories: (i) missing data, (ii) incorrect data, (iii) imprecise data, and (iv) irrelevant data. Examples of issues are missing events, missing case attributes, and incorrect timestamps. Mans et al. [19] use the taxonomy by Bose et al. [5] to evaluate data quality at the Maastricht University Medical Centre. Taking an interview-based approach, they assign an occurrence frequency to each quality issue, with the three most frequently occurring issues being missing events, imprecise timestamps, and imprecise resource information. In the same line of research, Mans et al. [20] discuss how data quality issues influence the potential of process mining to answer frequently asked questions by healthcare practitioners. They mainly focus on timestamp-related data quality issues: incorrect timestamps and timestamps recorded at an insufficiently granular level.

While Bose et al. [5] focus on the identification of event log quality issues, Suriadi et al. [24] both specify 11 data quality issues based on their experience and describe semi-automatic methods to rectify them. The proposed fixes often require domain knowledge to e.g. specify a minimal activity ordering. Moreover, the provided solutions are confined by the boundaries of the event log as this is the only data source considered. For example: a common issue involves data inserted into the HIS using electronic forms, implying that all events recorded when submitting the form share the same timestamp. To tackle this issue, Suriadi et al. [24] suggest merging all these events into a single event. While this approach can be defended when the event log is the sole data source, it needs to be recognized that information can be lost for analysis purposes. Hence, existing event log improvement literature can be extended by considering the use of other sources of process-related information. This paper considers ILS data, which has not been considered for this purpose yet.

In recent years, ILS data has been used for process mining purposes. It is either (i) used directly to perform process mining on sequences of locations, or (ii) converted to an event log using domain knowledge to apply process mining afterwards. ILS data has e.g. been directly used for process mining to retrieve patients' movement processes [7], to mine the workflow of medical devices [16], and to study a surgical process [8]. In these papers, a process instance consists of a sequence of locations and not a sequence of activities. Both only coincide when each activity takes place in a dedicated location. While this might be reasonable for highly specialized hospital units, this assumption will often not hold, e.g. when several activities are executed in a box at the emergency department.

In an effort to link ILS data to activities, Senderovich et al. [23] aim to convert ILS data to an event log. To this end, the interaction concept is introduced, which expresses a period of time during which e.g. a patient and a staff member are simultaneously present at a location. The detected interactions are mapped to activity labels using an integer linear program in which domain knowledge is encoded [23].

Despite the recent uptake in the use of ILS data for process mining purposes, the integration of an event log with its accompanying ILS data has not been considered in literature. Nevertheless, as will be argued in Section 4, ILS data can be helpful to alleviate data quality issues associated to HIS data.

4 Using ILS data to tackle event log quality issues

This section outlines, from a conceptual angle, the opportunities that ILS data offers to tackle event log quality issues. ILS data contains location patterns of process participants such as patients, medical staff and potentially even medical equipment. It enables to determine e.g. when a patient visited the room in which MRI-scans are made, even when this is recorded in the HIS at another moment. Besides location patterns, co-locations between process participants can also be identified by matching location pattern. A co-location, consistent with an interaction in [23], is a period of time during which multiple process participants are present at the same location. Co-location patterns convey valuable information for event log improvement as it typically reflects the execution of an activity.

To structure the remainder of this section, the 27 event log quality problems identified in [5] are used. When tackling these issues, ILS data will provide more solid support for some of them compared to others. To this end, a distinction is made between level 1 (Section 4.1) and level 2 support (Section 4.2). Quality issues for which level 1 support is provided require extensive domain knowledge to solve. However, ILS data can generate useful insights to facilitate the consultation of domain experts. Level 2 support means that ILS data provides a stronger foundation to directly enrich the event log or correct data errors. However, this does not imply that domain knowledge becomes redundant. A last group of issues are not considered in a healthcare context, as discussed in Section 4.3.

4.1   Level 1 support from ILS data

Missing activity names. This quality problem refers to the absence of activity names for particular events. A missing activity name can be retrieved from ILS data by looking for similar location or co-location patterns. However, many activities might have e.g. the same co-location pattern such as the co-location between a patient and a nurse. Consequently, domain experts will play an important role in mapping an ILS pattern to the appropriate activity.

Missing timestamps. A timestamp is missing when it is absent for an event. When events are recorded automatically by a HIS, e.g. after a click action, a timestamp is automatically generated and is, hence, unlikely to be missing.

This is consistent with the case study in [19], indicating that this is a quality issue with low occurrence frequency. When a timestamp is absent, ILS data can complement domain knowledge in an effort to insert a proxy for the missing timestamp, e.g. when the activity should be executed at a particular location.

Incorrect cases. This quality issue reflects the presence of cases in the event log which are related to another process. When ILS data centers around patients of a particular process, cases included in the ILS data can be compared to the cases included in the event log. Even when this is not the case, patients visiting a different zone in the hospital could be incorrect cases. However, domain knowledge is required to define particular filtering rules.

Incorrect events. Incorrect events are events which are recorded in the HIS, but did not occur in reality. ILS data can support the detection of such events by checking whether e.g. a co-location between a patient and a resource took place at or around the event's timestamp. In case of an incorrect event, no such pattern should be found in ILS data, implying that the event should be deleted.

Incorrect activity names. An incorrect activity name occurs when the name of the activity is registered incorrectly when e.g. clicking a drop-down menu value or entering it manually. ILS data can be helpful by detecting inconsistencies between the activity label of an event and the location or co-location of a patient in ILS data. Such inconsistencies can serve as an input for domain expert consultation. It should be noted that Mans et al. [19] mark this as a low frequent issue.

Imprecise relationships. This quality issue occurs when events cannot be linked to a case because of the case definition that is used. When studying patient-related healthcare processes, a patient will often be considered as a case. This is confirmed by the case study in [19], where this issue did not occur. When multimorbidity prevails, similar events might be associated to different conditions that a patient suffers and it might not be clear which events relate to the process under study. When such a connection is absent, ILS data can be used to study the locations which are visited or the medical staff that is involved.

Imprecise activity names. Imprecise activity names are activity names which are defined rather coarsely, causing them to occur multiple times for a particular patient, even though they refer to different actions. ILS data can support the domain expert by conveying insights in potential differences in location or co-location patterns between several occurrences of the same activity name.

Irrelevant cases. Irrelevant cases are present when the event log contains cases which are not relevant for a particular analysis. ILS data can support judging whether a particular patient is relevant when e.g. movement patterns or sequences of visited locations play a role. Hence, ILS data can support filtering operations in close consultation with domain experts.

Irrelevant events. Irrelevant events are events which are not relevant for a particular analysis question. When actions occurring at a particular location are not deemed relevant, ILS data can support filtering operations. Similar to irrelevant cases, this will require close interaction with domain experts.

4.2   Level 2 support from ILS data

Missing events. Missing events are events that have not been recorded for a patient. This can occur e.g. when particular events need to be recorded in the HIS manually. For instance: intermediate checkups by a physician or a nurse might not be recorded in the patient's file. This is, according to the case study in [19], one of the most frequently occurring data quality issues. ILS data can be used to detect those missing events as they will e.g. generate a co-location pattern between a patient and medical staff. Based on contextual information from domain experts, missing events can be imputed in the event log.

When only a single event, e.g. the start of a treatment, is recorded for each activity execution in the HIS, the corresponding complete event can also be seen as a missing event. In this respect, ILS data is a rich source of information to add events with other transaction types related to a particular activity execution. This does not only hold for start and complete events, but also for e.g. suspend and resume events defined by the XES lifecycle extension [12].

Missing/incorrect relationships. Missing and incorrect relationships are events which are not associated to a patient or associated to a wrong patient, respectively. When the activity under consideration requires a particular location or co-location pattern, ILS can be used to determine the associated patient or to rectify incorrect relationships. For instance: when co-location is required, it can be determined whether the resource associated to the event is co-located with a patient at a particular point in time. In the case study of Mans et al. [19], both missing and incorrect relationships were marked as a low frequency issue. The fact that missing relationships are infrequent can be attributed to the fact that all actions in a HIS are typically related to a specific patient.

Missing case attributes. A case attribute such as a patient's physical condition is missing when its value is not recorded for particular patients. ILS data is unlikely to support the specification of e.g. the patient's weight. However, it can be used to impute new location-related case attributes in the event log, which can be considered as missing case attributes from the event log perspective. Examples are the distance traveled or the number of locations visited.

Missing event attributes. When an event attribute value is absent, the missing event attributes issue occurs. Similar to missing case attributes, ILS data will probably not enable the specification of attributes which are completely unrelated to the patient's location. However, a location attribute can be added to the event log. Adding this information enables studying the use of particular hospital areas, determining the relationship between activities and locations, etc.

Incorrect/imprecise case/event attributes. This quality problem occurs when a wrong or inaccurate value for a case/event attribute is entered. For location-related attributes, ILS data can be leveraged to e.g. correct values which are recorded manually in the HIS or to provide a more detailed value. In [19], imprecise case/event attributes are absent and the occurrence frequency of incorrect case/event attributes is marked as low.

Incorrect timestamps. A recorded timestamp is incorrect when it does not correspond with the actual time of activity execution. Even though it is marked as a low frequent issue in [19], it should not be ignored as making registrations in the HIS is sometimes postponed. For instance: a physician might record his/her findings after visiting several patients. When such behavior is present, recorded timestamps will not coincide with actual activity execution. ILS data can be used to correct these timestamps as activity execution will be characterized by particular location or co-location patterns. For instance: a checkup by a physician is characterized by a co-location between a patient and a physician.

Imprecise timestamps. An imprecise timestamp is not recorded at a sufficiently detailed level but, e.g., at the date level. This is marked as a relatively frequently occurring issue in [19]. Similar to incorrect timestamps, ILS data can be leveraged to impute more detailed timestamps in the event log.

Missing resources. This quality issue implies that resource information is not recorded for a particular event. ILS data can be used to retrieve missing resource information by detecting a co-location of the patient associated to the event and a resource at that particular point in time. However, activity execution does not, by definition, require a co-location between a patient and a resource (e.g. when fulfilling an administrative task). When no co-location is required for an activity and it is known at which location it is executed, ILS data can still be helpful.

Incorrect resources. Incorrect resources imply that the resource associated to an event is not the one actually executing the activity. Even though it is indicated as non-occurring in [19], it can be quite common in healthcare when, e.g., all registrations on a particular computer take place under the account of one staff member. In this respect, ILS data can be used to determine which staff member was co-located with the patient at the moment the activity is executed.

Imprecise resources. Resource information is imprecise when it does not refer to a specific staff member, but e.g. to a staff category such as nurse or physician. It is one of the more frequent quality issues in the case study of Mans et al. [19]. ILS data can be used to impute more detailed resource information in the event log by detecting the execution of the activity in terms of a location or co-location pattern. When multiple staff members are involved, the HIS will probably only record the resource entering the activity in the system. In that sense, resource information can still be imprecise, even when it refers to a specific staff member. ILS data is highly relevant here as the co-location between multiple staff members indicates that several resources are responsible for activity execution.

4.3   Other event log quality issues

Missing/incorrect/imprecise position. Data quality issues related to the event's position in a trace are not taken into consideration. This is due to the fact that they relate to event logs without timestamps, which is not considered relevant within the context of a HIS.

Missing cases. Missing cases refer to patients for which no data is recorded in the HIS. As no file is recorded for these patients, they are not registered upon arrival. When ILS data is recorded by e.g. integrating an RFID tag in a patient identification wristband, it is likely that no ILS data will be recorded for these patients. However, this quality issue seems to be less relevant in healthcare, as is also supported by the case study in [19].

5   Challenges

From Section 4, it follows that ILS data can play an important role in improving the quality of a healthcare event log. However, to operationalize this data integration, several challenges need to be taken into account. In this section, four challenges are discussed, demonstrating the need for future research.

5.1   Presence of data quality issues in ILS data

While this paper focuses on event log quality issues, it should be recognized that ILS data can also suffer from data quality issues. Gal et al. [10] discuss ILS data quality challenges in queue mining, which is a subfield of process mining. In particular, they highlight the absence of a case identifier for some instances, the difficulty to determine the start and end point of activity execution, and issues related to reaching an appropriate level of data granularity for the analysis.

It should be noted that Gal et al. [10] and other research using ILS data for process mining use ILS data in isolation. This paper advocates the use of both ILS and HIS data. Consequently, HIS data can also be used to contextualize patterns observed in ILS data, which can be helpful to e.g. determine a missing case identifier. Nevertheless, data quality assessment of ILS data is still required prior to its use to alleviate event log quality issues. Data quality should also be a prime concern when the ILS is installed e.g. by performing data accuracy and data completeness tests [25]. Moreover, technology provider's middleware often automatically filters out some inaccuracies present in the data [13].

5.2   Simultaneous presence of event log quality issues

Section 4 outlined how ILS data can be helpful to alleviate a series of event log quality issues. In doing so, the perspective of one specific data quality problem is taken. However, in reality, multiple issues can be present simultaneously. Consider, for instance, that a nurse checks up on multiple patients and afterwards records it under the account of a colleague in the HIS. This constitutes a combination of the quality issues incorrect timestamps and incorrect resources.

To know which event log issues are present, systematic data quality assessment needs to be performed. Event log quality assessment is currently often carried out on an ad-hoc basis. Hence, developing and implementing a systematic and generic way to perform data quality assessment on an event log is an important research challenge. The data quality assessment tool should follow a

'signaling' approach in which potential issues are highlighted. Whether these latter issues actually constitute data quality problems requires domain knowledge as this might be context-dependent.

5.3   Need for a systematic way to capture domain knowledge

From the prior challenge and Section 4, the importance of domain knowledge becomes apparent. In order to use ILS data to improve healthcare event log quality, a relationship must be established between events from HIS data and location/co-location patterns in ILS data. Domain experts play a critical role in defining this relationship given the wide diversity of healthcare processes and HIS implementations. Consequently, there is a need for a systematic way to capture domain knowledge, marking an important research challenge.

To operationalize this, the basic idea of activity patters, introduced in [17] to map low-level events to high-level activities, can be leveraged. An activity pattern is a labeled process model containing the events registered during activity execution, and e.g. conditions related to resource use and timing restrictions.

Activity patterns can also be used to specify the relationship between HIS data and location/co-location patterns in ILS data. This implies that a set of intuitive activity pattern building blocks (with executable semantics) needs to be developed, relating to both HIS and ILS data. In contrast to [17], this enables hospital data specialists to create activity patterns themselves. A simplified example of an activity pattern for activity 'Follow-up patient treatment' is provided in Figure 1. It shows that the execution of the activity involves an ordering of an event (in the event log) and one or more co-locations between a patient and a nurse (in ILS data). Moreover, the constraint indicates that the follow-up of patient treatments can only take place in rooms C1 to C4. Future work will define more complex building blocks to e.g. express choice or optional event registration.
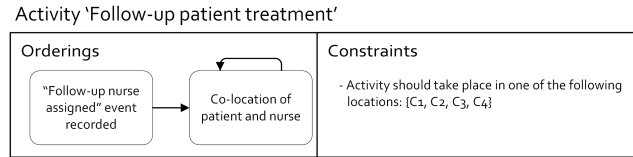

Fig. 1: Illustration of an activity pattern.

5.4   Need to perform the integration in a semi-automated way

The integration of HIS data and ILS data should be conducted in a semi-automated way. Using domain knowledge, captured in the form of activity patterns, an enhanced event log should be automatically created in which event log quality issues are tackled and ILS patterns are contextualized. During the data integration process, the data specialist can be asked for additional inputs when issues appear. This close interaction with the data specialist and the domain expert ensures that context-specific information is taken into account.

While shaping the semi-automated integration process already poses a research challenge, a related challenge is that the resulting enhanced event log will be location-aware. As the current XES-format [12] does not explicitly include location-related information, a novel location XES extension needs to be defined. This enables conducting location-aware process analyses, which will become more important in process mining given the increasing interest in ILS.

6   Conclusion

This paper discussed, from a conceptual angle, the opportunities that ILS data provides to tackle event log quality issues. This is a novel perspective as prior work on event log quality improvement did not consider the use of other sources of process-related data. ILS data can play an important role to alleviate quality issues considered important in healthcare such as incorrect/imprecise timestamps and imprecise resource information. While prior process mining research centers around the use of either HIS data or ILS data, this paper showed the benefits of integrating both to obtain an enhanced event log. As outlined above, ILS data can be used to tackle common event log quality issues. Conversely, the event log provides rich contextualization of patterns observed in ILS data.

Besides the potential benefits of ILS data to create an enhanced event log, this paper also outlined some challenges. Hence, further research is still required to systematically integrate HIS and ILS data. However, these efforts are worthwhile as a richer and more accurate enhanced event log will make an important contribution towards exploiting the full potential of process mining in practice. Moreover, the enhanced event log will enable the development of new techniques related to e.g. resource behavior analysis and patient waiting time analysis.

References

1. van der Aalst, W.M.P.: Process mining: data science in action. Springer, Heidelberg (2016)
2. van der Aalst, W.M.P., Adriansyah, A., ..., Wynn, M.: Process mining manifesto. Lecture Notes in Business Information Processing **99**, 169–194 (2012)
3. Batini, C., Scannapieco, M.: Data quality: concepts, methodologies and techniques. Springer, Heidelberg (2006)
4. Bendavid, Y.: RTLS in hospitals: technologies and applications. In: Encyclopedia of E-Commerce Development, Implementation, and Management, pp. 1868–1883. IGI Global, Hershey (2016)
5. Bose, R.J.C.P., Mans, R.S., van der Aalst, W.M.P.: Wanna improve process mining results? It's high time we consider data quality issues seriously. Tech. Rep. BPM Center Report BPM-13-02 (2013)
6. Caron, F., Vanthienen, J., Vanhaecht, K., Van Limbergen, E., De Weerdt, J., Baesens, B.: Monitoring care processes in the gynecologic oncology department. Computers in Biology and Medicine **44**, 88–96 (2014)
7. Fernandez-Llatas, C., Benedi, J.M., Garcia-Gomez, J.M., Traver, V.: Process mining for individualized behavior modeling using wireless tracking in nursing homes. Sensors **13**(11), 15434–15451 (2013)

8. Fernandez-Llatas, C., Lizondo, A., Monton, E., Benedi, J.M., Traver, V.: Process mining methodology for health process tracking using real-time indoor location systems. Sensors **15**(12), 29821–29840 (2015)

9. Forsberg, D., Rosipko, B., Sunshine, J.L.: Analyzing PACS usage patterns by means of process mining: steps toward a more detailed workflow analysis in radiology. Journal of Digital Imaging **29**(1), 47–58 (2016)

10. Gal, A., Senderovich, A., Weidlich, M.: Challenge paper: data quality issues in queue mining. Journal of Data and Information Quality **9**(4), paper 18 (2018)

11. Gschwandtner, T., Grtner, J., Aigner, W., Miksch, S.: A taxonomy of dirty time-oriented data. Lecture Notes in Computer Science **7465**, 58–72 (2012)

12. Gunther, C.W., Verbeek, H.M.W.: XES standard definition. Technical report, Eindhoven Unversity of Technology, Eindhoven, The Netherlands (2014)

13. Jeffery, S.R., Garofalakis, M., Franklin, M.J.: Adaptive cleaning for RFID data streams. In: Proceedings of the 32nd International Conference on Very Large Databases. pp. 163–174 (2006)

14. Kim, W., Choi, B.J., Hong, E.K., Kim, S.K., Lee, D.: A taxonomy of dirty data. Data Mining and Knowledge Discovery **7**(1), 81–99 (2003)

15. Kirchner, K., Herzberg, N., Rogge-Solti, A., Weske, M.: Embedding conformance checking in a process intelligence system in hospital environments. Lecture Notes in Computer Science **7738**, 126–139 (2013)

16. Liu, C., Ge, Y., Xiong, H., Xiao, K., Geng, W., Perkins, M.: Proactive workflow modeling by stochastic processes with application to healthcare operation and management. In: Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining. pp. 1593–1602 (2014)

17. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: Guided process discovery - a pattern-based approach. Information Systems **76**, 1–18 (2018)

18. Mans, R., Reijers, H., van Genuchten, M., Wismeijer, D.: Mining processes in dentistry. In: Proceedings of the 2nd ACM International Health Informatics Symposium. pp. 379–388 (2012)

19. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B.: Process mining in healthcare: evaluating and exploiting operational healthcare processes. Springer, Heidelberg (2015)

20. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B., Moleman, A.J.: Process mining in healthcare: data challenges when answering frequently posed questions. Lecture Notes in Computer Science **7738**, 140–153 (2013)

21. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. Journal of Biomedical Informatics **61**, 224–236 (2016)

22. Rovani, M., Maggi, F.M., de Leoni, M., van der Aalst, W.M.P.: Declarative process mining in healthcare. Expert Systems with Applications **42**(23), 9236–9251 (2015)

23. Senderovich, A., Rogge-Solti, A., Gal, A., Mendling, J., Mandelbaum, A.: The ROAD from sensor data to process instances via interaction mining. Lecture Notes in Computer Science **9694**, 257–273 (2016)

24. Suriadi, S., Andrews, R., ter Hofstede, A.H., Wynn, M.T.: Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Information Systems **64**, 132–150 (2017)

25. van der Togt, R., Bakker, P.J., Jaspers, M.W.: A framework for performance and data quality assessment of Radio Frequency IDentification (RFID) systems in health care settings. Journal of Biomedical Informatics **44**(2), 372–383 (2011)