



Defining Data Quality Issues in Process Mining with IoT Data

Yannis Bertrand^(✉), Rafaël Van Belle^(D), Jochen De Weerd^(D),
and Estefanía Serral^(D)

Research Centre for Information Systems Engineering (LIRIS), KU Leuven,
Warmoesberg 26, 1000 Brussels, Belgium
{yannis.bertrand,rafael.vanbelle,jochen.deweerd,
estefania.serralasensio}@kuleuven.be

Abstract. IoT devices supporting business processes (BPs) in sectors like manufacturing, logistics or healthcare collect data on the execution of the processes. In the last years, there has been a growing awareness of the opportunity to use the data these devices generate for process mining (PM) by deriving an event log from a sensor log via event abstraction techniques. However, IoT data are often affected by data quality issues (e.g., noise, outliers) which, if not addressed at the preprocessing stage, will be amplified by event abstraction and result in quality issues in the event log (e.g., incorrect events), greatly hampering PM results. In this paper, we review the literature on PM with IoT data to find the most frequent data quality issues mentioned in the literature. Based on this, we then derive six patterns of poor sensor data quality that cause event log quality issues and propose solutions to avoid or solve them.

Keywords: Data quality · Process mining · IoT data

1 Introduction

As IoT devices, i.e., sensors and actuators, are becoming increasingly more important for supporting the execution of business processes (BPs), there is a growing awareness of the opportunity to use the data collected by these devices for process mining (PM). Such IoT data can serve as a source for the derivation of an event log of the process around which IoT devices are placed, which can then be used to apply PM techniques (e.g., discovery, conformance checking).

However, IoT data (in particular sensor data) are well-known to be of poor general quality, i.e., suffering from noise, containing missing data, etc. There is a risk that underlying sensor data quality issues lead to data quality issues in the event log extracted from them, e.g., erroneous activity names, missing events, imprecise event-case relationships, etc.

Previous research has identified various event log quality issues [3] and patterns leading to some of those issues [27]. This being said, no work to date has

This research was supported by the Flemish Fund for Scientific Research (FWO) with grant number G0B6922N.

studied how the intrinsic characteristics of sensor data lead to event log quality issues and which specific patterns characterise event log quality issues stemming from quality issues in the source sensor data. This is of interest for research as identifying and understanding these patterns makes it easier for other researchers and practitioners to improve their IoT data quality to prevent event log quality problems and ultimately improve PM results.

In this paper, we address this gap and investigate data quality issues in PM that make use of IoT data. To do so, we review papers from the literature on IoT PM that mention data quality issues, both in sensor data and in the event logs derived from the sensor data. Based on this, we identify patterns of event log quality issues caused by quality issues in the source IoT data.

The remainder of the paper is structured as follows. In Sect. 2, we first go over the literature on data quality in general, before mentioning data quality in PM and IoT and outlining PM using IoT data. Then, in Sect. 3, we introduce our research questions and detail the methodology we followed to review the literature on data quality in IoT PM and derive patterns from it. After this, in Sect. 4, we present the results of our literature review and the patterns found in the literature. The results and the patterns are discussed in Sect. 5. We conclude our paper with suggestions to improve the quality of sensor data in IoT PM and ideas for future work.

2 Background

2.1 Data Quality

Data quality is a vast research topic and many definitions of data quality exist. In general, data quality is seen as the extent to which data meet the requirements of their users [25,30]. Various dimensions have been defined to describe and quantify data quality, among which: accuracy, timeliness, precision, completeness, reliability, and error recovery [16]. Note that the importance of each of these dimensions depends on the use case and the type of data.

2.2 Data Quality in Process Mining

Process mining assumes as input an event log consisting of all the events that took place in the process that is being analysed within a certain time frame. In order to apply process mining, an event log should include at least the following data elements: a case ID, indicating to which instance of the process an event belongs; a timestamp; and the label of the activity performed [24].

Data quality issues in PM revolve around errors, inconsistencies and missing data in event logs. The authors of [3] propose to classify these issues along two axes: the type of issue (incorrect, irrelevant, imprecise or missing data) and the event log entity affected (case, event, event-case relationship, case attribute, position, activity name, timestamp, resource, and event attribute). Some issues affecting events, timestamps and activity names are argued to be more important and are therefore analysed in further detail.

In [27], the authors build upon this framework and identify 11 event log quality issues in the form of imperfection patterns. For each of these patterns, a usual cause is identified, an example is given, a link is made with a event log quality issue from [3], and advice to detect and solve the issue is provided.

However, both seminal works focus on data quality issues arising in traditional event logs, while process mining on IoT data is faced with event log quality issues stemming from intrinsic characteristics and limitations of IoT devices.

2.3 IoT Data Quality

IoT data quality is a broad topic ranging from detecting IoT data quality issues to improving data quality through cleaning methods [16,28]. IoT applications often rely on low-cost sensors with limited battery and processing power, frequently deployed in hostile environments [28]. This leads to sensor issues such as low sensing accuracy, calibration loss, sensor failures, improper device placement, range limit and data package loss. Such sensor faults, in turn, cause various types of errors in the generated data complicating further analysis.

The authors of [28] reviewed the sensor data quality literature and listed the following error types (in decreasing order of frequency): outliers; missing data; bias; drift; noise; constant value; uncertainty; stuck-at-zero. When left untreated these errors result in incorrect data, and subsequent analysis will yield unreliable results, ultimately leading to wrong decisions.

To prevent misguided decision making, it is important to assess the underlying data quality. To this end, the authors of [21] introduced measures for sensor data quality: completeness, timeliness, plausibility, artificiality and concordance.

2.4 Process Mining with IoT Data

IoT devices usually sense the environment and produce at runtime a sequence of measurements called a *sensor log*, usually in the form shown in Table 1.

Table 1. Example of a sensor log generated by in smart spaces.

Timestamp	Sensor	Value
...
2022-05-31 12:34:52	M3	ON
2022-05-31 12:34:58	M5	OFF
2022-05-31 12:35:04	M3	OFF
2022-05-31 12:35:22	T2	22
2022-05-31 12:38:17	M29	OFF
...

The vast majority of the process mining literature involving IoT data focuses on deriving an event log from a sensor log. Traditional process mining techniques

can then be applied to this event log to, e.g., discover control-flow models of the processes. Typical steps include preprocessing the raw data (i.e., cleaning, formatting), event correlation to retrieve the cases each event belongs to and event abstraction to derive meaningful process events from sensor data (see, e.g., [5, 15, 18, 26, 29]).

These papers often report errors in the event logs derived from sensor data, which cause issues in the PM results (e.g., spaghetti models due to irrelevant events). In this paper, we argue that a large portion of the errors in the event log are due to data quality problems in the source sensor log, which are amplified by the event abstraction step and result in errors in the event log used for PM.

3 Methodology

In this section, we detail the methodology followed to review the literature on PM with IoT data and to derive patterns from the literature. It consists of three main steps: research question definition, literature selection and data extraction.

3.1 Research Questions

Three research questions (RQs) are addressed in this research:

- RQ-1: Which IoT data quality issues do IoT process mining papers face?
- RQ-2: Which event log quality issues do IoT process mining papers face?
- RQ-3: Which patterns can be found between IoT data and event log quality issues in IoT process mining?

3.2 Literature Selection

To answer these RQs, we scanned the literature on IoT PM that mentioned IoT data and event log quality issues. To do so, we devised a query consisting of three parts: process mining keywords, IoT data keywords and data quality keywords. After some refinements, the following query was finally selected:

(“process mining” OR “process discovery” OR “process enhancement” OR “conformance checking”) AND (“sensor data” OR “iot data” OR “internet of things data” OR “low-level log” OR “low-level data”) AND (“data quality” OR “data challenges” OR “data issues” OR “data preparation” OR “data challenge” OR “data issue”)

The query was executed on the Scopus and Limo online search engines, which access articles published by Springer, IEEE, Elsevier, Sage, ACM, MDPI, CEUR-WS and IOS Press. Because the literature tackling data quality in PM with IoT data is still very scarce, all fields were searched, yielding 177 results in total.

After removing duplicates and non-English results, papers were scanned based on title and abstract, before a full paper scan was performed. Papers were included based on their ability to answer the RQs, i.e., they had to apply PM with sensor data and mention data quality issues in sensor data or event

logs derived from sensor data or both. Review papers that could answer RQs were usually very generic and for this reason were excluded and replaced with the original studies, which answered the RQs in more detail. At the end of the review process, 17 studies remained for analysis (see Fig. 1 for more detail).

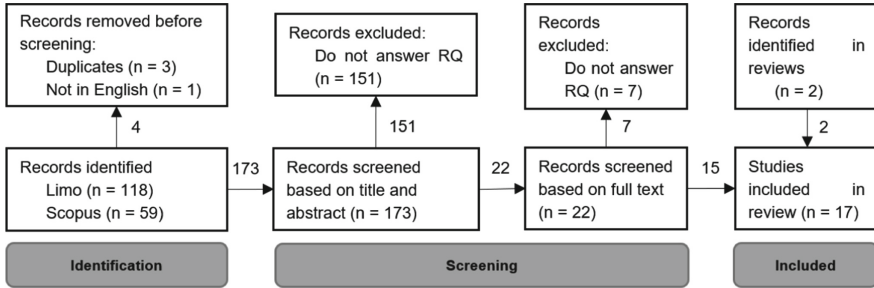


Fig. 1. Literature selection: included and excluded papers.

3.3 Data Extraction

The following information was extracted from the studies: The environment; The types of IoT data used and whether process data (i.e., a traditional event log) were also available; the IoT data and event log quality issues, following the classifications of [3, 28], respectively; and the analytical goal of the study (i.e., the type of PM to apply).

Based on this, patterns linking IoT data quality issues with event log quality issues were derived. For each pattern, its origin (cause of IoT data quality issue), effects (resulting event log quality issues) and potential remedies are discussed.

4 Results

4.1 Mapping of Data Quality Issues in IoT PM

The results of the data extraction can be found in Table 2. As can be seen, most of the papers report on process mining conducted in an industrial or healthcare environment. The vast majority of the literature uses only sensor data, from which an event log is derived (occasionally, mined models are shown in the papers), as discussed in Sect. 2.4. In line with the two most frequent environments considered by the papers, two main types of sensor data emerge: individual location sensor (ILS) data in healthcare and time series (TS) and discrete sensor data in industrial scenarios. These different data types are often affected by different data quality issues, which are discussed in the next paragraph. Finally, a slight upward trend can be seen in the number of publications over time, with a peak in 2018.

Concerning data quality issues, the most frequent IoT data quality issues encountered (RQ1) are noise (7), outliers (4) and missing data (4). Next to this,

many papers also mention volume (5) as a sensor data issue, which does not make the data erroneous, but can make the data considerably more difficult to analyse. Regarding event log quality issues (RQ2), the most frequent is incorrect event (7), followed by missing event (3), incorrect activity name (2) and incorrect

Table 2. Summary of the information extracted from the literature.

ID	Environment	Data type(s)	Data quality issue(s)		Goal	Year	Ref
			IoT data	Event log			
S1	Healthcare	Process data, TS sensor data	Outliers, noise	Incorrect events, irrelevant events	Process discovery	2012	[17]
S2	Healthcare	Individual location sensor (ILS) data	Outliers, noise	/	Process discovery	2013	[11]
S3	Logistics, healthcare	Process data, TS sensor data (simulated)	/	Missing event-case relationship, missing event attributes	Decision mining	2014	[9]
S4	Industry	TS sensor data	Volume, variety, velocity	/	Event log creation	2016	[20]
S5	Healthcare	ILS data	Inaccurate, granularity	Incorrect activity names, incorrect event-case relationship, missing events	Process redesign	2016	[31]
S6	Commerce	ILS data	/	Imprecise event-case relationship	Process discovery	2017	[14]
S7	Industry	GPS data	Missing data, volume	/	Predictive process monitoring	2018	[1]
S8	Industry	TS sensor data, discrete sensor data	Outliers, noise, duplicates	Incorrect event-case relationship	Event log creation	2018	[5]
S9	Industry	TS sensor data, discrete sensor data	Volume, granularity	/	Event log creation	2018	[6]
S10	Healthcare	ILS data	Volume	Incorrect events, incorrect timestamps	Queue mining	2018	[12]
S11	Healthcare	Hospital information system (HIS) and ILS data	Missing data, inaccurate data, granularity	Incorrect events	Event log repair	2018	[23]
S12	Home	ILS data	Noise	Incorrect events	Habit mining	2019	[8]
S13	Industry	TS sensor data, discrete sensor data	Outliers, noise	/	Event log creation	2020	[4]
S14	Industry	TS sensor data	Noise	Incorrect events, duplicate events	Anomaly detection	2020	[22]
S15	Various	Process data, sensor data	Noise, missing data, volume, granularity	/	Event log creation	2021	[2]
S16	Healthcare	ILS data	Noise, missing data	Missing events, incorrect events, duplicate events	Process discovery	2021	[10]
S17	Home	Video camera data	Noise	Incorrect events, missing events, incorrect activity names	Event log creation	2022	[19]

event-case relationship (2). Note that slashes in Table 2 indicate that the paper did not report data quality issues, which does not necessarily mean that no issue was encountered in the study.

4.2 Patterns Description

In this section, we present the patterns we have derived from the literature (RQ3). Note that papers that mention only either IoT data or event log quality issues cannot be used to derive patterns and, in addition to this, S11 cannot be used because the IoT data and event log quality issues described are unrelated (the event log is not derived from the IoT data). For each pattern, we discuss its origin, effects and potential remedies. Table 3 provides an overview.

Pattern 1: Incorrect Event-Case Relationship Due to Noisy Sensor Data. In many cases, when trying to derive an event log from sensor data, one of the main issues is that no case ID is present in the sensor log (e.g., in S8, S14, S17). To solve this problem, an event correlation step has to be performed, which will annotate events derived from the sensor log with the ID of the case they relate to. This correlation can be done either based on domain knowledge or using data-driven techniques. However, as noted in S8, this step is highly sensitive to the quality of the sensor data. In particular, noise and outliers can lead data-driven techniques to split cases mistakenly, resulting in labelling events with incorrect case IDs.

To avoid this issue, the use of sensor data cleaning methods is very important. The authors of S8 recommend in their follow-up paper S14 to use robust quadratic regression to clean and smoothen noisy sensor data.

Pattern 2: Erroneous Events Due to Inaccurate Location Sensor. ILS data is often used for PM, the assumption that different activities take place in different locations enabling a straightforward conversion of the sensor log into an event log (see, e.g., [13]). However, when different activities are executed in adjacent locations, there is a risk that several sensors will register the passage of a user (e.g., a patient, a resource) simultaneously. This generates erroneous events in the sensor log, which hinder the event abstraction step and result in incorrect events and activity names in the event log. This can have important consequences on PM: S16 reports that less than 0.5% errors in the event log already have a considerable impact on the quality of the process models mined.

This issue can best be treated by improving the sensor infrastructure. Using more accurate sensors or placing them further from each other can help avoid the issue completely. Otherwise, ex-post treatment can be applied by, e.g., deleting passages that last less than a given threshold (e.g., one minute in [7], cited by S12; 24 s in S5).

Pattern 3: Missing Events Due to Sampling Rate. Inadequate sampling rates can cause missing events in the event logs. It arises when the sampling rate of the sensors is too low, hence events that should be detected by these sensors are

not. In S5, for instance, the sampling rate of the system is 12 s, which means that passages of less than 12 s through a given location might not be recorded (which is realistic when the location is, e.g., a corridor), resulting in missing events.

The authors of S16 also propose a post-hoc solution to impute missing events based on the characteristics of the physical process environment. For example: given rooms A, B and C, if C is only accessible via B, then a user must have been through B even if the sensor log only contains passages in A and C. Other possibilities involve improving sensor logging a priori by fine-tuning the sampling rate for each location (so there are neither missing events nor incorrect events), e.g., lowering the sampling rate of the sensor in the corridor while increasing the sampling rate of the sensor in the doctor's practice. A second possibility is to filter out passages that are too short (e.g., in S5, passages of less than 24 s are considered as noise and removed). This technique can be refined by using a low sampling rate in all locations and filtering out events that are obviously too short or too long, depending on the location.

Pattern 4: Missing Events Due to Sensor Range Limit. A similar pattern arises in the dimension of space rather than time. In this case, the range of sensors (e.g., location sensors) is too narrow and does not encompass the whole area where an activity could take place. This issue leads to missing events that happened beyond the sensor's reach. For instance, in S5, the range of location sensors is two meters, which means that any movement beyond this range will remain unnoticed, hence if an activity of the process is executed more than two meters from the sensor, it will not be detected.

The post-hoc solution suggested by S16 (see Pattern 3) can be applied to impute missing events that are caused by a lack in sensor range. In addition to this, improving the coverage of the physical process space by installing additional sensors can help prevent this issue from happening.

Pattern 5: Erroneous Events Due to Noisy Sensor Data. In this pattern, noise is present in the sensor data due to issues during logging or due to the presence of noise affecting the phenomenon measured by the sensors (e.g., in S17, video data contain sequences that are irrelevant for the process). This noise in the sensor data is picked up in the event abstraction phase and translates into noise in the event log in the form of incorrect events and events that carry incorrect activity name.

To solve this issue, S17 uses the inductive miner - infrequent (IMf) discovery algorithm, which has a parameter that can be adjusted to determine the level of infrequent behaviour to include in the model mined. The same approach is followed by S14, also using the noise threshold of the IMf algorithm to determine which events to leave out of the model.

Pattern 6: Incorrect Timestamps Due to Sensor Range Limit. This issue is related with P4, and arises when the arrival of a user in a room/at a location does not coincide with the beginning of the activity executed here. This causes the beginning of the activity to be recorded earlier than the actual beginning of

the activity. E.g., in S10, it is assumed that the beginning of a consultation with a doctor is the moment when the patient is detected by the location sensor in the office of the doctor. However, as noted by the authors, it may be that the doctor is still busy in another room, or finishing taking notes for the previous patient. The same issue can also affect the end of the activity, when a user leaves the room with a certain delay after the end of the activity.

This issue can sometimes be solved by modifying the placement of the sensors, to make them detect users more precisely when events happen, or by adapting the range of the sensors to make them only detect users when the activity actually started or ended (and not after it started or before it ended either).

Table 3. Overview of identified patterns linking sensor faults and data quality issues to the associated errors in process mining.

	Sensor fault/characteristic \implies Sensor data issue \implies Process mining errors	References
P1	Unstable environment \implies Noisy sensor data & outliers \implies Incorrect case ID	S8, S14, S17
P2	Inaccurate sensor location \implies Duplicate or inconsistent sensor readings \implies Incorrect events, incorrect activity names	S16
P3	Inadequate sensor sampling rate \implies Events not captured in sensor log \implies Missing events	S5
P4	Sensor range limit \implies Activities outside sensor range are missing \implies Missing events	S5, S16
P5	Unstable environment \implies Noisy sensor data \implies Incorrect events, incorrect activity names	S1, S12 S14, S17
P6	Sensor range limit \implies Activity start/end time is logged incorrectly \implies Incorrect timestamps	S10

5 Discussion

It is interesting to note that the most frequently IoT data quality issues are among the most cited error types in the IoT literature. However, the high number of papers mentioning noise as an issue and the absence of other, more refined, IoT data quality issues from [28] makes us suspect that some of the papers reviewed used noise and outliers as bucket terms for more specific sensor data quality issues (e.g., drift, bias). This may have also had an effect on the precision of the patterns we found.

Next to this, it is remarkable that the patterns identified usually result in issues with the most critical event log elements (i.e., event, case ID, activity name, timestamp). This is mainly due to the fact that PM using IoT data often focuses on extracting these required elements from the sensor log. Moreover, these elements being the most essential also makes it more likely that errors concerning them are searched for (and detected). This effect can also be observed in [27], where most patterns detected cause event log quality issues affecting events, activity names or timestamps.

The literature mentions two main strategies to improve sensor data quality: post-hoc data cleaning (e.g., removing outliers, smoothing; for a complete discussion of sensor data cleaning techniques, see [28]) and fostering good data logging practices (e.g., careful sensor data placement, constant environmental conditions). While the latter has the advantage of preventing the issue rather than solving it, it must be noted that completely preventing sensor data quality issues is impossible. E.g., sensor failure is typically hard to detect, let alone predict [16]. Moreover, some of the patterns are interrelated, and avoiding one of them sometimes comes at the cost of aggravating another one. For instance: ILS can only avoid blind spots (Pattern 4) at the cost of having zones where multiple location sensors overlap (Pattern 2). This means that some data cleaning will always have to be performed, e.g., to impute missing events due to blind spots in between location sensors.

Finally, it is worth noting that some papers use sensor data to repair traditional event logs collected by information systems. S11, for instance, uses ILS data to detect sequences of events that are not realistic given the path followed by patients in a hospital and correct them. S11 also argues that neither sensor data nor event logs collected by traditional sources are fully reliable, and that the main advantage of using two (or more) data sources is to be able to compare them to find anomalous data and hopefully correct them.

6 Conclusion

In this paper, we investigated data quality issues in PM using IoT data. After reviewing background literature and related works on sensor data quality and event log quality, we scanned the literature to find the most common sensor data quality issues (RQ1) and event log quality issues (RQ2) in IoT PM papers, following well-established data quality taxonomies [3,28]. Based on this, we identified six patterns of sensor data quality issues that cause event log quality issues and hinder IoT PM (RQ3), and mentioned possible remedies to the underlying IoT issues.

Following this, our advice for improving sensor data quality for PM is to first improve the logging practices, with 1) thoughtful sensor placement to avoid missing and duplicate events; 2) use devices to identify the users tracked by IoT devices to have case IDs at logging time; 3) careful choice of sensors to obtain data at the best granularity level (i.e., accuracy, frequency) to avoid huge volumes of data. Second, we encourage researchers to investigate more generic

and more automated techniques (i.e., requiring little expert input) to detect and correct sensor data quality issues, as data cleaning approaches mentioned in the literature are often ad-hoc and highly tailored for data from specific sensors. Finally, we align ourselves with [23] in advising researchers and practitioners to try to combine different data sources whenever possible.

One key limitation of this study is the fact that we restricted ourselves to patterns that could be derived from the existing IoT PM literature. Accordingly, given the still fairly low maturity of this subdomain of PM, we cannot make founded claims on completeness of these patterns. In particular, with IoT PM focusing heavily on the derivation of events and subsequent control-flows from sensor data, there is a lack of research into using IoT data for non-control-flow related data, including event and case attributes, e.g., in function of decision mining, trace clustering, etc. Such uses of IoT data are very likely to produce additional data quality patterns. Another important area for future research concerns the streaming nature of typical IoT data, given the additional complexity this creates for data quality detection and rectification strategies. Finally, while well-known as a data quality issue in the field of IoT, the level of measurement precision of sensors is currently not yet taken into account within the IoT PM literature. Given the importance of delicately tuned thresholding approaches, e.g. for event abstraction, we consider research on the impact of sensor data precision on process mining results to be another promising area for future work.

References

1. Bandis, E., Petridis, M., Kapetanakis, S.: Business process workflow mining using machine learning techniques for the rail transport industry. In: Bramer, M., Petridis, M. (eds.) SGAI 2018. LNCS (LNAI), vol. 11311, pp. 446–451. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04191-5_37
2. Beverungen, D., et al.: Seven paradoxes of business process management in a hyper-connected world. *BISE* **63**(2), 145–156 (2021)
3. Bose, R.J.C., Mans, R.S., van der Aalst, W.M.: Wanna improve process mining results? In: 2013 IEEE CIDM, pp. 127–134. IEEE (2013)
4. Brzywczy, E., Gackowiec, P., Liebetrau, M.: Data analytic approaches for mining process improvement-machinery utilization use case. *Resources* **9**(2), 17 (2020)
5. Brzywczy, E., Trzcionkowska, A.: Creation of an event log from a low-level machinery monitoring system for process mining purposes. In: Yin, H., Camacho, D., Novais, P., Tallón-Ballesteros, A.J. (eds.) IDEAL 2018. LNCS, vol. 11315, pp. 54–63. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03496-2_7
6. Brzywczy, E., Trzcionkowska, A.: Process-oriented approach for analysis of sensor data from longwall monitoring system. In: Burduk, A., Chlebus, E., Nowakowski, T., Tubis, A. (eds.) ISPEM 2018. AISC, vol. 835, pp. 611–621. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97490-3_58
7. Dogan, O., Bayo-Monton, J.L., Fernandez-Llatas, C., Oztaysi, B.: Analyzing of gender behaviors from paths using process mining: a shopping mall application. *Sensors* **19**(3), 557 (2019)
8. Dogan, O., et al.: Individual behavior modeling with sensors using process mining. *Electronics* **8**(7), 766 (2019)

9. Dunkl, R., Rinderle-Ma, S., Grossmann, W., Anton Fröschl, K.: A method for analyzing time series data in process mining: application and extension of decision point analysis. In: Nurcan, S., Pimenidis, E. (eds.) CAiSE 2014. LNBIP, vol. 204, pp. 68–84. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19270-3_5
10. Fernandez-Llatas, C., Benedi, J.M., Gama, J.M., Sepulveda, M., Rojas, E., Vera, S., Traver, V.: Interactive process mining in surgery with real time location systems: interactive trace correction. In: Fernandez-Llatas, C. (ed.) Interactive Process Mining in Healthcare. HI, pp. 181–202. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-53993-1_11
11. Fernández-Llatas, C., Benedi, J.M., García-Gómez, J.M., Traver, V.: Process mining for individualized behavior modeling using wireless tracking in nursing homes. *Sensors* **13**(11), 15434–15451 (2013)
12. Gal, A., Senderovich, A., Weidlich, M.: Challenge paper: data quality issues in queue mining. *JDIQ* **9**(4), 1–5 (2018)
13. Grefen, P., Brouns, N., Ludwig, H., Serral, E.: Co-location specification for iot-aware collaborative business processes. In: CAiSE 2019, pp. 120–132. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-21297-1_11
14. Hwang, I., Jang, Y.J.: Process mining to discover shoppers' pathways at a fashion retail store using a wifi-base indoor positioning system. *IEEE T-ASE* **14**(4), 1786–1792 (2017)
15. Janssen, D., Mannhardt, F., Koschmider, A., van Zelst, S.J.: Process model discovery from sensor event data. In: Leemans, S., Leopold, H. (eds.) ICPM 2020. LNBIP, vol. 406, pp. 69–81. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72693-5_6
16. Karkouch, A., Mousannif, H., Al Moatassime, H., Noel, T.: Data quality in internet of things: a state-of-the-art survey. *JNCA* **73**, 57–81 (2016)
17. Kaymak, U., Mans, R., Van de Steeg, T., Dierks, M.: On process mining in health care. In: SMC 2012, pp. 1859–1864. IEEE (2012)
18. Koschmider, A., Janssen, D., Mannhardt, F.: Framework for process discovery from sensor data. In: EMISA, pp. 32–38 (2020)
19. Kratsch, W., König, F., Röglinger, M.: Shedding light on blind spots-developing a reference architecture to leverage video data for process mining. *DSS* **158**, 113794 (2022)
20. Krumeich, J., Werth, D., Loos, P.: Prescriptive control of business processes. *BISE* **58**(4), 261–280 (2016)
21. Kuemper, D., Iggena, T., Toenjes, R., Pulvermueller, E.: Valid. iot: a framework for sensor data quality analysis and interpolation. In: ACM MMSys 2018, pp. 294–303 (2018)
22. Maeyens, J., Vorstermans, A., Verbeke, M.: Process mining on machine event logs for profiling abnormal behaviour and root cause analysis. *Ann. Telecommun.* **75**(9), 563–572 (2020)
23. Martin, N.: Using indoor location system data to enhance the quality of healthcare event logs: opportunities and challenges. In: Daniel, F., Sheng, Q.Z., Motahari, H. (eds.) BPM 2018. LNBIP, vol. 342, pp. 226–238. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11641-5_18
24. Reinkemeyer, L.: Process mining in a nutshell. In: Process Mining in Action, pp. 3–10. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-40172-6_1
25. Scannapieco, M.: Data Quality: Concepts, Methodologies and Techniques. Data-Centric Systems and Applications. Springer, Heidelberg (2006). <https://doi.org/10.1007/3-540-33173-5>

26. Seiger, R., Zerbato, F., Burattin, A., Garcia-Banuelos, L., Weber, B.: Towards iot-driven process event log generation for conformance checking in smart factories. In: EDOCW, pp. 20–26. IEEE (2020)
27. 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. *ISJ* **64**, 132–150 (2017)
28. Teh, H.Y., Kempa-Liehr, A.W., Wang, K.I.K.: Sensor data quality: a systematic review. *J. Big Data* **7**(1), 1–49 (2020)
29. Valencia Parra, Á., Ramos Gutiérrez, B., Varela Vaca, Á.J., Gómez López, M.T., García Bernal, A.: Enabling process mining in aircraft manufactures: extracting event logs and discovering processes from complex data. In: *BPM2019IF* (2019)
30. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. *JMIS* **12**(4), 5–33 (1996)
31. Zhang, Y., Martikainen, O., Saikkonen, R., Soisalon-Soininen, E.: Extracting service process models from location data. In: Ceravolo, P., Guetl, C., Rinderle-Ma, S. (eds.) *SIMPDA 2016. LNBIP*, vol. 307, pp. 78–96. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-74161-1_5

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

