Information & Management 61 (2024) 103973

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

Demystifying data governance for process mining: Insights from a Delphi study

Kanika Goel [a,*], Niels Martin [b,c], Arthur ter Hofstede [a]

[a] School of Information Systems, Queensland University of Technology, Brisbane, Australia
[b] Hasselt University, Research group Business Informatics, Hasselt, Belgium
[c] UHasselt - Expertise Centre for Digital Media, Diepenbeek, Belgium

ARTICLE INFO

Keywords:
Data governance
Process mining
Delphi study
Process data

ABSTRACT

Data governance is recognised as a new capability for organisations to maximize the value of data. Process mining is essential for the resilient growth of businesses, making process data a strategic asset for organisations. Even though the availability of reliable process data is vital for obtaining dependable insights into process mining techniques, there exists no framework that explains how to govern process data holistically. We address this gap by presenting the first data governance framework for process mining that was derived from a Delphi study conducted with a panel of academics and practitioners from around the world. The framework provides multiple avenues for future research.

1. Introduction

Data is recognised as fundamental for analysing and improving organisational performance. There is growing awareness of its role in strategic decision making [1,2], which has brought the significance of *data governance* to the fore. Data governance has been defined as the exercise of authority and control over the management of data [3]. It aims to build a corporate-wide strategy to maximise the use of data within organisations and reduce data-related risks [4,5].

Data has gained increased significance in enterprises and government institutions [6], especially as its complexity and volume continued to explode. According to Statista, the total volume is expected to grow exponentially from 33 zettabytes in 2018 to 2100 zettabytes in 2035 [7]. This challenge has required organisations to develop more sophisticated capabilities regarding data use, for which different ways to combine, manipulate, and store data and convert it to valuable information are needed [8].

The need for data governance is magnified by the growing data regulations and standards that organisations need to comply with. For example, the General Data Protection Regulations (GDPR) legislation has created pressure on businesses to keep a strong hold on what data is collected, where it is stored, and how it is used [1]. Furthermore, in a study by Holt et al. [9], 45 % of participants drawn from a global community of database and data professionals did not have data governance policies and practices in place. Data governance in information systems is an area of emerging interest; however, despite its high importance, it remains an underresearched area and is infrequently practised in industry [8,10,11].

For organisations, process data is becoming an increasingly important type of data, providing a valuable asset to aid the improvement of organisational processes [12]. Process data refers to data regarding the execution of organisational processes that is automatically recorded by the information systems that support these processes [12,13]. We consider process data as data recorded by an information system that will be used as input for analysis algorithms whose insights are aimed at better understanding the operational execution of an organisational process.

Process data can be leveraged to generate an event log, an analysable dataset that captures the sequence of activities performed in an organisational process for a particular case (i.e., process instance), when these activities were executed, and potentially other details, such as who executed them. Process mining, a specialised form of data-driven process analytics, then can be used to uncover the real-life behaviour of business operations from an event log [12]. A myriad of process mining techniques have been developed in industry and academia [14] to help management understand the real-life flow of organisational processes, determine the conformance of processes with a normative model, and enhance processes [12]. Process mining has been applied in multiple

* Corresponding author.
E-mail address: kan.goel03@gmail.com (K. Goel).

https://doi.org/10.1016/j.im.2024.103973
Received 6 September 2022; Received in revised form 17 April 2024; Accepted 6 May 2024
Available online 7 May 2024

industries and has been recommended for driving the resilient growth of businesses [15]; for example, a major oil and gas company applied process mining to achieve better insight into its internationally centralised invoice reconciliation process and identified some major inefficiencies. After improving these inefficiencies, the organisation saved 1.2 million euros [15].

To ensure that the insights from the use of process mining techniques are dependable, the availability of reliable process data is vital. Compared to datasets that serve as input for typical data mining applications, process data in the form of an event log has specific characteristics that pose unique data-related challenges. For example, events are related to one another due to the notion of a case present in process data, which groups together events related to one specific process instance. Moreover, each event is characterised by an activity label reflecting the activity to which the event is linked. A notion of order among events also must be present, which is typically derived from the timestamp associated with each event [12]. This shows that process data has particularities that must be taken into consideration. To leverage the great potential that process mining offers, process data needs to be treated as a strategic asset, highlighting the significance of data governance for this type of data.

The unique characteristics of process data have led to dedicated research efforts on specific data governance topics, including process data quality [16–18], process data privacy [19,20], and event log construction [21,22]. These lines of research are leading to the development of novel approaches, taking the specific context of process data as a starting point. While these contributions help move the knowledge base forward on specific data governance subjects, our review of the process mining literature revealed that a more holistic set of data governance considerations for process data has not been identified. Against this background, this paper investigates the following research question: *What are the data governance considerations for process data?*

To approach this research question, we performed a Delphi study with an international panel of academics and practitioners who are experts in process mining and data governance. A Delphi study is a structured process of collecting and distilling knowledge from a panel of experts in multiple rounds using questionnaires with controlled feedback [23]. Our Delphi study resulted in a list of 38 data governance considerations for process data. These data governance considerations are structured around 11 dimensions, including the 10 dimensions proposed by the Data Management Association (DAMA), a global association dedicated to advancing the concepts and practices of information and data management.

This study contributes to the areas of process mining and data governance, demonstrating the critical importance of explicitly considering process data in data governance strategies and policies. Along with demonstrating how process data can be transformed into a strategic asset, we also identify important avenues for future research at the intersection between process mining and data governance.

The remainder of this paper is structured as follows. Section 2 provides the relevant background on data governance and process mining. Section 3 presents an outline of the adopted research method and study design. Section 4 details the findings of the study, and Section 5 presents a discussion around them. The paper ends with a conclusion in Section 6.

2. Background

2.1. Data governance

Data governance is defined as the "planning, oversight, and control over management of data and data-related resources" [3]. Data governance provides a cross-functional framework, which allows managing data as a strategic asset. In doing so, it formalises data policies, standards, and procedures and also monitors compliance [1]. Data governance is different from data management. According to DAMA, "data management is the development, execution, and supervision of plans, policies, programs, and practices that control, protect, deliver, and enhance the value of data and information assets" [3], whereas data governance is the exercise of control and authority over the management of data assets [3]. Therefore, data governance complements data management. Data governance establishes policies and procedures around data, whereas data management enacts those policies and procedures to manage data. Data governance aims to promote a culture that fosters the development of corporate-wide data policies, guidelines, and standards that are consistent with the vision, mission, and objectives of the organisation [8]. Overall, data governance includes the organisation of data management, ensuring alignment with business needs, compliance, and a common understanding of data [24].

The main driver of data governance is the increasing significance of digital transformation for the growth of companies, and thus the increased consideration of data as an asset [25]. As a consequence, data governance is assuming a position of greater importance in organisations [1,25]. Data governance has been identified as a new "business capability to obtain value from data" [2]. Many organisations have created dedicated teams or a dedicated department around data collection, integration, and analysis [26]. Such a dedicated team looks after data innovation and governance, which is then used for strategy implementation [27,28].

Despite its significance, data governance remains an under-researched area [1,8,9] that has gained attention only recently. For example, Zhang et al. [11] presented a strategic action framework to deploy data governance in organisations, Al-Badi et al. [29] proposed a data governance framework for big data, AlRuithe et al. [30] promulgated a data governance framework for cloud computing, and Begg and Caira [10] described a data governance framework for small and medium enterprises. In 2019, Abraham et al. [1] conducted a rigorous systematic literature review as a foundation for a conceptual framework for data governance consisting of six dimensions that provide an understanding of the key areas that must be considered for data governance. These prior research efforts illustrate the significance of research in data governance.

An external body widely cited in the area of data governance, DAMA International is "a not-for-profit, vendor-independent global association of technical and business professionals dedicated to advancing the concepts and practices of information and data management" [31]. The purpose of DAMA is to promote the understanding, development, and practice of managing data and information as key enterprise assets to support organisations. DAMA has defined ten dimensions that should be considered with respect to data governance [3]: data architecture, data modeling and design, data storage and operations, data security, data integration and interoperability, documents and control, reference and master data, data warehousing and business intelligence, meta-data, and data quality (Table 1). In this paper, we use the DAMA framework as a structure for understanding data governance, given its wide recognition in the data governance field.

2.2. Process mining

Process mining aims to extract knowledge from process data to understand, improve, and monitor business processes [12]. A business process is commonly defined as "a set of logically related tasks performed to achieve a defined business outcome" [32]. Today, processes are typically supported by information systems, which record data about the execution of a process [12,13]. For use as an input for process mining, process data is structured as an event log. An event log consists of cases that represent process instances, for example, a specific loan application in a loan application process. For each case, a series of events is recorded while the process is being executed. Each event is described by a number of attributes, such as the activity linked to the event, the time at which the event occurred, and possibly other attributes, such as the resource associated with the event [12].

The event log serves as an input for a wide variety of process mining

Table 1
DAMA dimensions and definitions adapted from Mosley et al. [3].

DAMA Dimension	Definition
Data architecture	Definition of the overall structure of data and data-related resources as an integral part of the enterprise architecture
Data modeling and design	Analysis, design, implementation, deployment, and maintenance of data solutions to maximise the value of the data resources to the enterprise.
Data storage and operations	Development, maintenance, and support of structured data to maximise the value of the data resources to the enterprise. Data operations management includes two subfunctions: database support and data technology management.
Data security	Planning, development, and execution of security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets.
Reference and master data management	Ongoing reconciliation and maintenance of reference data and master data. Reference data include defined domain values, standardised terms, code values, business definitions, and more. Master data consist of data associated with business entities.
Data integration and interoperability	Acquisition, extraction, transformation, movement, delivery, replication, federation, virtualisation, and operational support of data.
Data warehousing and business intelligence management	Collection, integration, and presentation of data to knowledge workers for the purpose of business analysis and decision making
Document and content management	Control over the capture, storage, access, and use of data and information stored outside relational databases
Metadata management	Control over the set of processes that ensure proper creation, storage, integration, and control to support associated use of meta-data
Data quality management	Definition, monitoring, and maintenance of data integrity and improvement of data quality

techniques. Recently, van der Aalst [33] distinguished six types of process mining: process discovery, conformance checking, performance analysis, comparative process mining, predictive process mining, and action-oriented process mining. *Process discovery* algorithms are used to retrieve a process model from an event log [33,34]. These process models typically represent the control flow—the order of activities in a process [34]—but also can represent, for instance, social networks depicting how staff members interact in a process [35,36]. *Conformance checking*, in its turn, aims to evaluate whether the behaviour described in the event log aligns with the behaviour displayed by a given normative process model [33,37].

Although process discovery and conformance checking are two very prominent areas of process mining, the four other process mining types described by van der Aalst [33] demonstrate the broad scope of the process mining field. *Performance analysis* focuses on the use of an event log to understand the performance of a process, for example, in terms of the service time, waiting time, or level of rework [33,38]. *Comparative process mining* leverages event logs to make intraorganizational or interorganisational process-related comparisons [33,39,40].

While the foregoing process mining types tend to be backward-looking as they focus on conveying an understanding of the process using historical data, *predictive process mining* involves making process-related predictions starting from an event log and thus constitutes a branch of process mining that is forward-looking [33,41]. *Action-oriented process mining* aims to convert diagnostics into actions by providing triggers to improve the execution of the process (e.g., by escalating a case when checks are skipped) or by identifying tasks that can be automated [33,42].

The literature points to several challenges when working with process data. For example, the quality of input data is critical for reliable process mining insights, as inaccurate or incorrect data can generate misleading insights and unjustified decisions [43,44]. Thus, there is a need for rules and guidelines to govern the quality of process data. Privacy and security of data is another crucial factor [20,45] as failing to comply with rules and regulations can have negative ramifications for the organisation. Furthermore, having the right data model is important to enable process mining on data stored in databases [46]. Access to correct metadata or annotations is another important factor to support the reliability of process mining insights [43]. In short, research in the areas of data quality, privacy, data modeling, and metadata has brought the unique requirements of process data to the fore, demonstrating the need to give specific consideration to process data.

To obtain reliable process mining insights, process data should be governed appropriately. As indicated by Zuboff [47], IT can have a transformative effect to *automate* as well as to *informate*. Algorithms and machinery allow IT to perform complex computational tasks, referred as the power to *automate*. Equally important is to *informate* the use of IT to capture the right data, which can then be used to improve business operations [48]. Possessing the competence of having appropriate data that can be used for improvements is the main objective of data governance.

Despite the significance of process data governance, it remains an underresearched area [8,25]. Recently, Goel et al. [16] proposed ImperoPD, a data governance framework for managing data quality for process mining. This framework highlights five business areas—business strategy management, organisation and process management, process data management, information technology management, and people management—and 20 business capabilities to consider when managing data quality for process mining. Although the framework provides interesting insights, it focuses solely on data quality, which is only one of many dimensions of data governance. van Cruchten and Weigand [21] have presented an event log management framework to improve the quality of event logs, structured around three areas: event log governance, event log lifecycle, and event log foundations. van Cruchten and Weigand [21] focused on security and privacy, metadata, and quality aspects of the event log, which constitutes three of the ten DAMA dimensions. This paper extends this stream of literature, taking a more holistic perspective on data governance by not restricting itself to selected data governance dimensions. This broad perspective is essential because process data has its own unique requirements, which need to be understood and harnessed for organisations to be able to improve their business operations. Furthermore, in this paper we gather insights from academics and practitioners from the fields of process mining and data governance, while the work of Goel et al. [16] and van Cruchten and Weigand [21] is based solely on the literature.

3. Research method and study design

3.1. Delphi study as a research method

Against the background of the limited knowledge on process data governance, a Delphi study was conducted to identify the relevant data governance considerations for process data and to assess their comparative relevance. Delphi studies are widely used in the areas of Information Systems (IS) and Business Process Management (BPM) to gain consensus among a group of experts on a topic of interest through multiple rounds of questionnaires [49–51]. IS scholars have been using the Delphi method for almost four decades, and its use has contributed to the field through its unique method for accessing the knowledge of experts [23]. The Delphi method has been applied in IS and BPM research studies to, for example, identify the constituent values of a BPM supportive cultural setting [52], identify the reasons for process deviance [53], identify performance indicators to benchmark hospital information systems [54], and examine opportunities and challenges for process mining in organisations [55].

A Delphi study is exploratory in nature and follows a structured process involving multiple iterations in which expert opinions are collected and distilled using a series of questionnaires [23,49]. Experts remain anonymous, to avoid any bias from confrontation [56]. In each round of the Delphi study, participants communicate their opinions through a questionnaire, which is collected and edited by the research team to provide the participants with a statement on the position of the panel, as well as each participant's own position [23]. Therefore, a Delphi study is an inductive, data-driven approach that is well suited for studies with research questions for which limited or no evidence exists [50].

3.2. Study design

This study was designed to identify data governance considerations for process data, as well as their comparative relevance. The design was shared with the expert panel, who were encouraged to comment on it. This section outlines the five key characteristics of the study design.

First, this study used a rating-type Delphi study design, which is a variant of the commonly used ranking-type Delphi study [23,56]. Based on the work of Schmidt [57], the study included three stages: brainstorming, narrowing down, and rating [23,57]. The brainstorming phase established an initial list of data governance considerations for process data (round 1), which was presented to the panel again for validation (round 2). In the narrowing down phase (round 3), panel members were asked to select the considerations that they deemed relevant, to reduce the list to a manageable size for the rating phase [23]. In the rating phase (rounds 4–6), panel members were asked to rate all considerations by assigning each to one of the predefined relevance categories. Hence, while the third stage of a rating-type Delphi requires *rating* all considerations, a raking-type Delphi study would have required experts to *rank* all considerations (i.e., assign all of them to ordered ranks). Given the multifaceted character of data governance, we anticipated receiving a substantial number of considerations. Because ranking is possible for only a limited number of items given the cognitive load on the experts, rating was considered a more suitable approach than ranking [53,55].

Second, to ensure a common understanding of data governance, we provided the expert panel with the ten functional areas for data governance from DAMA as background information [31]. A definition (adopted from DAMA) was provided for each of the dimensions. It should be stressed that the DAMA dimensions were provided merely as background information. To keep an open view, we purposefully refrained from constraining the collection of data governance considerations around the DAMA dimensions during the first round by not mandating that the expert panel stick to them. The question in the first round was very open, as the experts were requested to identify data governance considerations that they perceived as "either specific to process data or have a specific interpretation for process data, impacting the value that an organisation can draw from process mining." The experts were not asked to provide the considerations according to different DAMA dimensions. Providing the DAMA dimensions as background information is in accordance with Paré et al. [23]. Kobus and Westner [58] stressed the significance of providing clear instructions to the experts, which to us implies that the meaning of key terminology, such as data governance, is clear to the experts.

Third, based on the input of the first round, the data governance dimensions of DAMA were deemed useful for categorising the data governance considerations for process mining. This enabled us to present the intermediate results to the expert panel in an organised manner. We were not restricted by the DAMA dimensions, however; for considerations that did not fit into any of the dimensions proposed by DAMA, new dimensions could be created based on input provided by the experts.

Fourth, we invited experts from both academia and industry to ensure that we captured the diversity of the field [23]. We also recruited process mining and data governance experts from different countries [52]. We defined formal requirements for recruiting experts [56]; academics had to have a PhD and at least 3 years of relevant academic experience. They also had to have been involved in real-life projects related to process mining during their academic experience. Practitioners had to have at least 3 years of work experience in the field of process mining.

Finally, we collected qualitative and quantitative feedback to determine the quality and convergence of the results of the study. We measured the respondents' satisfaction with the overall study as well as with the coding of responses. This quantitative feedback is a well-established feature of Delphi studies [52,55]. To this end, a 7-point Likert scale was used: 1, extremely dissatisfied; 2, moderately dissatisfied; 3, slightly dissatisfied; 4, neither dissatisfied nor satisfied; 5, slightly satisfied; 6, moderately satisfied; and 7, extremely satisfied. These measures were used to assess the convergence of results. In addition, we also requested qualitative feedback regarding the coding performed and the study as a whole. Overall, we aimed for a high satisfaction score (mean >5.5 and standard deviation close to 1) and positive qualitative feedback [23]. Analysis of the satisfaction score also enabled us to check for selection bias, that is, to verify whether an improvement in satisfaction scores occurred because experts were more satisfied and not because dissatisfied respondents dropped out of the study [59].

3.3. Preparation activities

To ensure an appropriate level of expertise within the expert panel, we used the selection criteria defined in Section 3.2. A Delphi study involves multiple iterations and thus requires a high level of commitment from the respondents, which is why we decided to approach experts from our academic and professional networks. We first identified 58 candidates for the expert panel, including 29 practitioners and 29 academics. Of these candidates, 55 (28 practitioners and 27 academics) met our selection criteria. We contacted these 55 candidates and allowed them to nominate other candidates they deemed suitable for this study. Twenty-six candidates agreed to participate in the study; however, 5 of these 26 candidates did not participate in round 1, dropping the number of initial respondents to 21.

The panel was well balanced in terms of academics (11 out of 21 respondents) and practitioners (10 out of 21 respondents). The panel members had experience with both process mining and data governance. We had panel members from ten countries. Further investigation of the collected background information revealed that our panel members were balanced in terms of expertise in technical and business areas, as evidenced by their PhD discipline and job position. The experts had used process mining in a wide range of sectors, including healthcare, retail, manufacturing, education, and others. Details are provided in Appendix A, Table A.4.

To ensure the clarity of the first round questionnaire, we conducted a pilot test in accordance with the established quality criteria for Delphi studies [55,60]. The pilot aimed to ensure that the questionnaire was understandable, the wording was unambiguous, and an appropriate introduction to the topic areas being investigated was provided to the participants [60]. The questionnaire included the goal of the study, some background on process mining and data governance, and an introduction to the dimensions of data governance by DAMA [3]. We distributed the questionnaire to four PhD students in the IS field and asked them to complete it. All four students provided their input. We also held a follow-up meeting to discuss any issues the students faced while completing the questionnaire. The added value of the background information on data governance was stressed, to establish an understanding of this rather abstract concept. Based on feedback from the four students, we made some minor changes to the design of the questionnaire to further enhance its understandability.

Furthermore, we established a set of coding guidelines (in accordance with Saldaña [61]) for coding expert responses before starting the study. We used a hybrid coding approach [62] that combines both inductive and deductive coding. Inductive coding is a data analysis process in which a researcher interprets data to develop concepts [63, 64], whereas a deductive coding data analysis process uses an a priori template of codes to analyse data [62]. Because we used the DAMA dimensions to make sense of the collected input but were open to new dimensions, the hybrid coding approach was suitable for this study.

Coding proceeded in multiple iterations. In iteration 1, all the responses were coded as a theme. In iteration 2, all similar themes were grouped together into categories. In iteration 3, categories along with their themes were allocated to a dimension of DAMA if the match was considered suitable. If not, a new dimension was created to contain the category and the themes.

3.4. Delphi study procedure and key figures

Our Delphi study consisted of six rounds and lasted for five months. In each round, the experts had between two and four weeks to complete an online questionnaire. Furthermore, the experts received an option to provide qualitative feedback on the coding and how the study was progressing. In accordance with the quality criteria for a Delphi study [23,60], the questionnaire consisted of detailed instructions, an overview of responses from the previous rounds (from round 2 onward), and changes to the results from the previous round. Table 2 provides an overview of the number of participants, the number of included data governance considerations, and satisfaction scores for each round of the Delphi study. The remainder of this subsection briefly outlines each round of this study and provides key information on the operationalisation of these rounds. Section 4 provides an overview of the key findings of the study, i.e., the data governance considerations. Further details are provided in Appendix B.

In the brainstorming phase, we first collected an initial list of data governance considerations for process data (round 1). The responses were consolidated by the research team into 36 data governance considerations, which were categorised in 11 dimensions: the 10 DAMA dimensions and "*Supporting organisational policies*"). In the next round (round 2), the experts were asked to validate the data governance considerations from round 1, and encouraged to change, add to, or reassign the considerations.

This process resulted in an increase in the number of considerations from 36 to 50. These 50 considerations were the input for the narrowing down phase (round 3), which aimed to reduce the number of considerations to a manageable number for the rating phase [23]. To this end, experts were asked to select the consideration(s) they found relevant. We applied a straightforward majority rule, according to which only considerations supported by at least 50 % of the experts were retained. This resulted in a reduction of the number of data governance considerations from 50 to 38.

In the next three rounds, which constituted the rating phase (rounds 4–6), the experts rated the shortlisted considerations according to their comparative relevance. For this, the experts could select either "irrelevant," "slightly relevant," "moderately relevant," "relevant," or "strongly relevant."

The number of active panel members ranged between 17 (round 6) and 21 (round 1). The dropout rate was 19 %, within the normal range considering the degree of commitment required of experts in a Delphi study [65]. The dropout rate also can be attributed to our strict policy of not inviting participants for the next round if they failed to answer the questionnaire in the pervious round. We did this because we wanted the participants to be aware of all the communication occurring during the study as well as the outcomes of the previous round. Nonetheless, despite the dropouts, our final panel size was within the recommended range of sample size for a Delphi study [23,56,66].

Two measures of satisfaction can be distinguished: coding satisfaction and overall study satisfaction. The satisfaction with coding was calculated using the input to the question "How satisfied are you with the coding after the *first/second* round?' ,[1] which was also asked at the end of rounds 2, 3 and 4. The mean score of satisfaction with coding among panel members was high and increased over the rounds (from 5.62 out of 7 in round 2 to 6.17 in round 4). Moreover, the associated standard deviation decreased over the rounds. The overall satisfaction score for the study in turn was calculated using the input to the question "How satisfied are you with the study?," which was asked at the end of each round from round 2 onward. In each round, the mean and standard deviation for the satisfaction scores were calculated. The mean satisfaction score was close to 6 (out of 7) since the start of the study, and the associated standard deviation was around 1, which is what we hoped to achieve in this study. The slight drop in the overall study satisfaction in rounds 5 and 6 can be attributed to a limited number of experts who rated their satisfaction as one category lower. This might be attributed to a certain level of expert fatigue toward the end of the study [51], especially given that the rating distributions were fairly stable over the rating rounds. To rule out the risk that satisfaction scores were to some extent artificially inflated by dissatisfied experts leaving the study, we examined the satisfaction scores of experts in the round before they dropped out. Besides one moderately dissatisfied expert who left the panel after the second round, the other three experts who dropped out provided satisfaction scores of at least 5 out of 7 in terms of both coding and the overall study immediately before leaving the study. Overall, the high satisfaction values indicate that we reached a convergence after round 6 of the study.

4. Results

Table 3 presents the key results of the Delphi study, that is, the final data governance considerations resulting from the study. As discussed previously, the data governance considerations were grouped according to the DAMA dimensions proposed in Mosley et al. [3]. The grey highlighted rows in the table represent the data governance dimensions proposed by DAMA. All the dimensions proposed by DAMA are included in the results, as they were all deemed relevant for process data governance by the experts. Moreover, we have derived a new dimension, "*supporting organisational policies and programs,*" from our analysis of the results. This new dimension refers to organisational policies and programs that support data governance efforts related to process data.

Table 2
Overview of the Delphi study procedure and statistics.

Phase	Brainstorming		Narrowing down	Rating		
Round	1	2	3	4	5	6
Active panelists	21	21	20	18	17	17
Academics	10	10	9	7	7	7
Practitioners	11	11	11	11	10	10
Data governance considerations[a]	36	50	38	38	38	38
Satisfaction - Overall study[b]						
Mean	–	5.90	5.95	6.11	5.94	5.76
Standard deviation	–	1.14	0.94	0.90	1.14	1.15
Satisfaction - Coding[c]						
Mean	–	5.62	5.65	6.17	–	–
Standard deviation	–	1.13	1.18	0.86	–	–

[a]After coding or voting.
[b]Likert scale from 1 to 7 (not assessed before Round 2).
[c]Likert scale from 1 to 7 (only assessed from Round 2 until Round 4, reflects the satisfaction with the coding results of the previous rounds).

[1] The formulation with "first" was used in round 2 and with "second" used in round 3. In round 4, the question to assess coding satisfaction was "How satisfied are you with the final list of considerations?"

Table 3
Final list of data governance considerations for process data.

Dimension/ Consideration	Definition	R6 Rating Distribution					R6 Median	R6 Mode
		IRR	SLR	MR	R	STR		
Data Architecture								
Process-centric data architecture (C.1)	There is a need for a well-documented data architecture that is process focused, considering process mining from the start rather than an afterthought.	0.00 %	0.00 %	35.29 %	11.76 %	**52.94 %**	STR	STR
Data Modeling and Design								
Modeling process mining requirements while designing the data model (C.2)	There is a need for the data model to capture all fields required for process mining in a format that enables the flexible use of process mining tools.	0.00 %	0.00 %	5.88 %	**76.47 %**	17.65 %	R	R
Identify and implement appropriate integrity constraints for Process data (C.3)	There is a need for the data model to comprise the necessary integrity constraints to collect appropriate process data.	0.00 %	5.88 %	23.53 %	**58.82 %**	11.76 %	R	R
Proper understanding of the data model (C.4)	There is a need for an in-depth and up-to-date understanding of the data model in order to build an event log.	0.00 %	5.88 %	17.65 %	23.53 %	**52.94 %**	STR	STR
Support multiple level of abstractions (C.5)	There is a need for the data model to support multiple levels of abstractions for process data.	0.00 %	0.00 %	11.76 %	**82.35 %**	5.88 %	R	R
Data Storage and Operations								
Data storage guidelines (C.6)	There is a need to maintain data storage guidelines providing information on what process data to retain and for how long, also considering its implications at the level of information systems.	0.00 %	5.88 %	23.53 %	**58.82 %**	11.76 %	R	R
Widely Adopted cross-organisational event log storage standard (C.7)	There is a need for a widely adopted cross-organisational event log storage standard that enables unified data and metadata storage, as well as facilitated data sharing.	0.00 %	11.76 %	11.76 %	**76.47 %**	0.00 %	R	R
Data Security								
Privacy preservation versus analysis preservation (C.8)	There is a need to weigh the implementation of privacy-preserving techniques for human process participants against their impact on the process mining potential.	0.00 %	5.88 %	23.53 %	**47.06 %**	23.53 %	R	R
Privacy preservation of organisation's sensitive data in a cross-organisational context (C.9)	There is a need to ensure the privacy of an organisation's sensitive data when performing process mining in a cross-organisational context.	0.00 %	5.88 %	11.76 %	35.29 %	**47.06 %**	R	STR
Store process data in a secure environment (C.10)	There is a need to store process data in a physically safe environment with appropriate cybersecurity controls.	0.00 %	0.00 %	11.76 %	35.29 %	**52.94 %**	STR	STR
Data integration and interoperability								
Cross-system data integration (C.11)	There is a need to integrate data from more than one system or database to create proper event logs, which may require data conversion, mapping, and use of knowledge graph methods.	0.00 % 17	0.00 %	0.00 %	**76.47 %**	23.53 %	R	R
Collection of end-to-end process data (C.12)	There is a need to collect data for the end-to- end process instead of considering a fragment of the process.	0.00 %	0.00 %	**41.18 %**	17.65 %	**41.18 %**	R	MR, STR
Event log construction guidelines and methods (C.13)	There is a need for guidelines and methods regarding event log construction, covering both technical (e.g., the extraction of process data) and nontechnical (e.g., the involvement of both data and process experts) aspects.	0.00 %	0.00 %	23.53 %	**52.94 %**	23.53 %	R	R
Event log transformation guidelines and methods (C.14)	There is a need for guidelines and methods supporting repeatable event log transformation to contribute to the transparency and explainability of process mining analyses.	0.00 %	0.00 %	**41.18 %**	23.53 %	35.29 %	R	MR
Process data interoperability between organisational entities (C.15)	There is a need to ensure process data interoperability between departments or partnering organisations to conduct process mining within that scope.	0.00 %	0.00 %	47.06 %	**52.94 %**	0.00 %	R	R
Documents and Content								
Store a rich set of process-, user-, and domain-related artefacts in a process repository (C.16)	There is a need to leverage process data to support the storage of a rich set of process-, user-, and domain-related artefacts such as process documentation, process models, user profiles, and ontologies in a process repository.	0.00 %	11.76 %	**70.59 %**	17.65 %	0.00 %	MR	MR
Reference and Master Data								
Master data specification (C.17)	There is a need to identify and implement appropriate methods to specify master data for processes.	0.00 %	5.88 %	5.88 %	**82.35 %**	5.88 %	R	R
Up-to-date master and reference data (C.18)	There is a need to ensure that the master and reference data is up-to-date at all times.	0.00 %	5.88 %	0.00 %	41.18 %	**52.94 %**	STR	STR
Identify critical process data (C.19)	There is a need to identify critical master and reference data (e. g., set of activities, events, or values of attributes) to get maximal value from process data governance efforts.	0.00 %	5.88 %	11.76 %	**58.82 %**	23.53 %	R	R
Data Warehousing and Business Intelligence								
Data warehouse for process mining (C.20)	There is a need to maintain a data warehouse for process mining, potentially integrated in a data lake, ensuring that process data from various systems is available in a schema that facilitates usage of process mining tools.	0.00 %	0.00 %	**41.18 %**	23.53 %	35.29 %	MR	R
Availability of real-time process data (C.21)	There is a need to ensure the availability of real-time process data to support decision making.	0.00 %	17.65 %	**52.94 %**	23.53 %	5.88 %	MR	MR

(*continued on next page*)

Table 3 (*continued*)

Dimension/ Consideration	Definition	R6 Rating Distribution					R6 Median	R6 Mode
		IRR	SLR	MR	R	STR		
Build support for process mining (C.22)	There is a need to make dashboards and other (preferably multi-purpose) tools available to build organisational support for collecting, storing, and analysing process data.	0.00 %	0.00 %	5.88 %	**76.47 %**	17.65 %	R	R
Make insights from process data insights broadly available in the organisation (C.23)	There is a need to make understandable and reliable process data insights available to knowledge workers and decision makers to enable monitoring and analysis of processes as well as for decision making.	0.00 %	0.00 %	5.88 %	23.53 %	**70.59 %**	STR	STR
Maintain a repository of process mining artefacts (C.24)	There is a need to maintain a repository of process mining artefacts, which can be leveraged in future analyses.	0.00 %	5.88 %	5.88 %	**82.35 %**	5.88 %	R	R
Meta-data								
Repository of process-related meta-data (C.25)	There is a need to establish and maintain a complete, up-to-date, and easily accessible repository of process-related meta-data, which provides detailed information regarding the meaning of process data (e.g., a glossary providing the interpretation of all attributes and their values).	0.00 %	5.88 %	11.76 %	**52.94 %**	29.41 %	R	R
Document the mapping between real-life process concepts and data recording (C.26)	There is a need for careful documentation and a perfect understanding of all domain expertise regarding how real-life process concepts such as tasks are captured as process data.	0.00 %	5.88 %	5.88 %	**64.71 %**	23.53 %	R	R
Document process data provenance (C.27)	There is a need to carefully document data provenance to make the source and transformation of process data fully traceable.	0.00 %	5.88 %	23.53 %	**41.18 %**	29.41 %	R	R
Document internal controls related to process data (C.28)	There is a need to carefully document the internal controls that are in place related to process data.	0.00 %	5.88 %	17.65 %	**70.59 %**	5.88 %	R	R
Data Quality								
Awareness of specific data quality issues for process data (C.29)	There is a need for awareness for data quality issues that are specific to process data to allow the use of reliable and unambiguous data.	0.00 %	0.00 %	0.00 %	**70.59 %**	29.41 %	R	R
Data quality assessment guidelines and rules for process data (C.30)	There is a need to maintain and document guidelines and rules regarding the assessment of data quality issues in process data, taking into account the use case at hand.	0.00 %	0.00 %	5.88 %	**58.82 %**	11.76 %	R	R
Assure the quality of process data transformations C.31)	There is a need to assure the quality of process data transformations, given their great impact on process mining outcomes.	0.00 %	0.00 %	0.00 %	**58.82 %**	41.18 %	R	R
Build trust in the truthfulness of process data (C.32)	There is a need to build trust in the truthfulness of process data to gain the confidence of business users in process mining results.	0.00 %	0.00 %	11.76 %	**58.82 %**	29.41 %	R	R
Supporting Organisational Policies and Programs								
Ensure consistency of process data governing policies and programs (C.33)	There is a need to ensure the consistency of process data governance policies and programs across the different data governance dimensions due to their close interconnection.	0.00 %	5.88 %	17.65 %	**41.18 %**	35.29 %	R	R
Training and communication programs for human resources (C.34)	There is a need to develop training and communication programs for human resources to correctly use systems, enabling the collection of reliable process data.	0.00 %	11.76 %	11.76 %	**70.59 %**	5.88 %	R	R
Identify process- related information needs (C.35)	There is a need to identify the information needed to make business decisions regarding processes and align process data governance efforts with it.	0.00 %	5.88 %	5.88 %	17.65 %	**70.59 %**	STR	STR
Process data and metadata ownership (C.36)	There is a need to assign dedicated responsibilities within the organisation regarding process data and meta-data.	0.00 %	5.88 %	17.65 %	**58.82 %**	17.65 %	R	R
Explicitly link strategic priorities to process data (C.37)	There is a need to explicitly link the organization's strategic priorities to the process data that can be leveraged to support achieving them.	0.00 %	0.00 %	35.29 %	**58.82 %**	5.88 %	R	R
Centre of excellence for process data analysis (C.38)	There is a need to facilitate the development of a center of excellence on process data analysis within the organisation.	0.00 %	23.53 %	**58.82 %**	5.88 %	11.76 %	MR	MR

Each row in Table 3 lists a data governance consideration, its definition, its rating distribution[2] after round 6, its median rating for round 6, and its mode rating for round 6. To illustrate the interpretation of the table, the first row consists of the data governance consideration for the dimension *data architecture*, designated *process-centric data architecture*. The definition is "there is a need for a well-documented data architecture that is process-focused, considering process mining from the start rather than an afterthought." The next five columns show that none of the respondents found the consideration irrelevant or slightly relevant, 35.29 % of respondents found the consideration moderately relevant, 11.76 % of respondents found the consideration relevant, and the majority of the respondents (52.94 %) found the consideration strongly relevant. Both the median rating and mode rating for this consideration in round 6 were strongly relevant.

Table 3 clearly shows that, apart from six considerations, "relevant" or "strongly relevant" was chosen by most the experts for all other considerations. For the other six considerations, "moderately relevant" was chosen most frequently. None of the considerations was consider "irrelevant" or "slightly relevant," demonstrating the significance of

[2] Please note that IRR stands for Irrelevant, SLR stands for Slightly Relevant, MR stands for Moderately Relevant, R stands for Relevant, and STR stands for Strongly Relevant in Table 3.

each data governance consideration for process mining.

5. Discussion

This section elaborates on our findings. After a general discussion in Section 5.1, the implications of our study are outlined in Section 5.2. The section ends with an overview of the key limitations in Section 5.3.

5.1. General discussion

Process mining offers huge potential to support organisations in improving their processes [12,55]. Process data is the key input for all process mining techniques and needs to be governed as a strategic asset to reach the full potential of process mining for industry and society. Given the specific characteristics of process data, dedicated research has been conducted on specific topics, such as process data quality (e.g., Martin [17]; Supriadi et al. (2017)). Although such targeted research efforts are highly valuable and underscore the importance of giving specific attention to process data, a more holistic understanding of data governance for process data is lacking in the knowledge base. By identifying data governance considerations for process data, this paper contributes to such an understanding. The considerations are identified via a Delphi study in which an international panel of experts shared their views following a structured procedure, consisting of multiple rounds.

The key contribution of our study is a structured list of 38 data governance considerations for process data considered relevant by the expert panel. These considerations cover a wide variety of data governance topics and highlight the need to take a process data view of these matters when organisations want to leverage the full potential of process data to improve their processes. They support the need to treat process data as first-class citizens, an idea that previously advanced as a guiding principle for the practical use of process mining in the influential *Process Mining Manifesto* published in 2011 [67]. To this end, organisations should establish a consistent set of process data governance policies and programs (C.33) and delineate clear responsibilities for these matters within the organisation (C.36).

Various considerations that emerged from the study stress the need to consider process mining from the design stage, rather than as an afterthought. This relates to the formulation of a process-centric data architecture (C.1), as well as the design of a data model. Such a data model, which should be properly understood in the organisation (C.4), needs to capture all required data fields (C.2) and integrity constraints (C.3) relevant to process mining. This is line with prior reports, such as Jans et al. [22], de Morillas et al. (2020) and Schuh et al. [68], which echo that process mining has its own data requirements, and that it is essential that appropriate data are captured to ensure that the results are of value. The data model should be flexible to support the extraction of event logs at multiple levels of abstraction (C.5). To determine the critical process data that needs to be captured, a clear delineation of the process-related master and reference data is needed (C.19), which should be kept up to date (C.18). With respect to the specification of master data, the panel also added that appropriate methods should be in place to support this specification within the organisation (C.17).

Panel members also clearly communicated the need for guidelines outlining how to handle process data in a variety of data governance topics. This relates to data storage (C.6), event log construction (C.13), event log transformation (C.14), and data quality assessment (C.30). The need for guidance in event log construction and transformation corroborates findings reported by Martin et al. [55], where complex data preparation emerged as a highly relevant challenge for the use of process mining in organisations. The need for support in data quality assessment [69], along with understanding the root causes behind process data quality issues [70], have been highlighted in prior research. When developing guidelines, it is important to consider the distinguishing characteristics of process data, such as the specific data quality issues that can prevail (C.29). In addition to the development of guidelines, staff members also need to be trained such that they can correctly use the organization's systems to improve data registration at the source (C.34), the latter an ambition that has been echoed in the literature [17].

Another set of considerations focuses on facilitating the actual use of process data within the organisation. This has both organisational and technical angles. From an organisational angle, it is important to explicitly link the organization's strategic priorities to the relevant process data (C.37), as well as to clearly identify the process-related information needs of the organisation (C.35). From a technical perspective, a dedicated data warehouse for process mining (C.20), as well as the need to have process data available in real time (C.21), were proposed by the panel. For example, Vogelgesang and Appel Rath [71] proposed the concept of Cube, a data warehouse–based approach for multidimensional process mining. Once insights have been generated from the available process data, these can be distributed within the organisation by embedding them in dashboards or other tools (C.22). The significance of dashboards for conveying process mining insights has been demonstrated by Ibanez-Sanchez et al. [72]), who developed such tools as Papp to present process indicators and combine process-based perspectives with key performance indicators. In any case, process mining insights should be shared as broadly as possible within the organisation (C.23). Process mining artefacts also can be stored in a repository to facilitate their reuse (C.24), and this repository can also be connected to a broader process repository that also contains, e.g., process documentation and process models (C.16).

To move the use of process data forward, all relevant expertise on process data analysis within the organisation can be brought together in a center of excellence (C.38). The significance of a business process management center of excellence has been recognised in the literature [73,74]. Such a centre can provide guidance to organisations in building efficient processes that abide by industry standards. Interestingly, a centre of excellence is also proposed to maintain governance for process data in our Delphi study. A key premise to foster the use of process data is that business users have confidence in the truthfulness of the data (C.32), highlighting the need to be aware of specific data quality issues of process data (C.29), as well as to ensure the quality of the data transformation performed (C.31). The importance of data quality stated by the experts is in line with prior research in which researchers focused on identifying key data quality dimensions to consider when working with process data [17,75].

Although the preceding paragraphs have focused mainly on process data itself, the importance of the associated metadata also clearly emerges in the results. In particular, the expert panel highlighted the need to establish and maintain a repository of process-related metadata (C.25). Within this realm, the importance of documenting how real-life process concepts are captured in process data (C.26), and which internal controls are in place (C.28), also have been pointed out. Moreover, the provenance of process data must be carefully registered to ensure its traceability (C.27). When relating the latter to prior work, Goel et al. [43] stressed the significance of metadata and proposed maintaining annotations for transformation, provenance, and data quality purposes. The insights from our Delphi study convey additional details (e.g., process-related metadata, such as name of process, version, and more) that need to be maintained to use process data in an optimal manner.

Our results also demonstrate that data governance considerations should move beyond the boundaries of a single organisational unit, which is typically needed to get a view of end-to-end processes (C.12), and even beyond organisational boundaries. This highlights the need to integrate data from different systems (C.11) to ensure the interoperability of process data between departments and/or organisations (C.15) and to widely adopt a cross-organisational event log storage standard to facilitate data sharing (C.7). In a cross-organisational context, privacy preservation of sensitive data is an important consideration (C.9), even though the panel considers the trade-off between privacy preservation on the one hand and the process mining potential on the other hand as an important trade-off in general (C.8). The latter is mentioned as a specific consideration under data security, along with the obvious need to store process data in a secure environment (C.10). This attention to privacy is in accordance with research on privacy-preserving process mining, such as Liu et al. [76] and Pika et al. [20].

When assessing the list of data governance considerations as a whole, it should be noted that it contains considerations that are relevant for process data but are applicable to other types of data as well. This is related to, for example, the need to store process data in a safe environment (C.10) and to have process data available in real time (C.21). On the other hand, a wide variety of considerations specific to process data and process mining came to the fore. Examples include the need for event log construction guidelines and methods (C.13) and the maintenance of a repository of process mining artefacts (C.24). Although we do not claim that these latter items cannot be transposed to other types of data, they will require explicit attention within a process data context.

The 38 considerations conceptualised for data governance for process mining augment the data governance framework presented by DAMA [3]. As discussed in Section 2, the DAMA framework comprises 10 dimensions, but it does not present concrete subdimensions and metrics for each of the dimensions proposed as significant for data governance. Only high-level guidance in the form of different deliverables for each of the dimensions is provided. For example, the key deliverables for the data security dimension in the DAMA framework are data security policies, data privacy and confidentiality standards, user profiles, passwords and memberships, data security permissions, data access views, document classifications, authentication and access history, and data security audits [3]. In our study, as a part of the security dimension, we observed the relevance of storing process data in a secure environment, ensuring that privacy of organisational data is maintained in a cross-organisational context and maintaining privacy of data during analysis. Although some considerations may overlap (e.g., storing process data in a secure environment) with the DAMA deliverables, others are new within the context of process data. Taking the data quality dimension as another example, the DAMA deliverables are improved

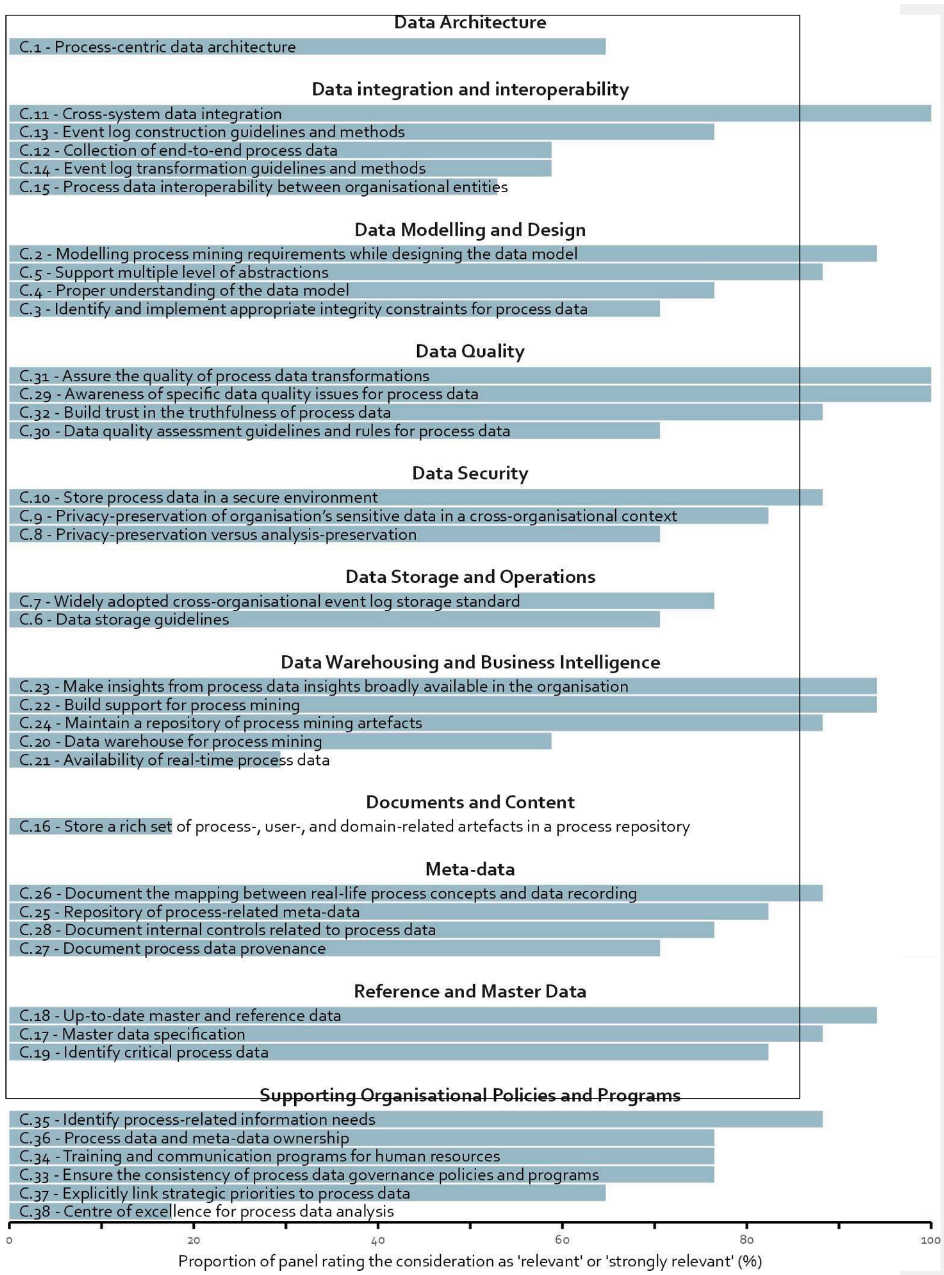

Fig. 1. Summary of the data governance considerations for process data.

data quality, data management and operational analysis, data profiles, data quality certification reports, and data quality service level agreements [3].

Our study presents new data quality considerations for process data, including guidelines for assessing process data quality and ensuring quality of process data transformations. Therefore, while our study uses the dimensions of the DAMA framework as a lens to structure the data governance field, the considerations obtained and presented here emerged from considering the specific context of process data. These considerations can be used to devise subdimensions and metrics for governing process data. Moreover, a novel dimension, "*supporting organisational policies and programs,*" has been added to cover the broad input provided by the experts. The inclusion of this dimension reflects the experts' view that organisations need to develop organisational policies and programs that support data governance efforts related to process data. For instance, the experts signaled the need to consider process data at the strategic level (C.37) and to explicitly identify process-related information needs (C.35).

Along with providing an overview of data governance considerations for process data, this paper also provides insight into their comparative relevance. Fig. 1 provides a visual summary of our 38 considerations and also highlights the comparative relevance of each consideration with shading. Each shade represents the proportion of the panel that rated the consideration as "relevant" or "strongly relevant" in the last Delphi round. All considerations from Table 3 are included in Fig. 1. This also holds for considerations with a relatively low proportion of the panel rating them as "relevant" or "strongly relevant" (e.g., C.16 and C.38), as they are still considered relevant by a majority of the panel. If not considered relevant by a majority of the panel, these considerations would have been removed in the narrowing down phase (round 3 of the Delphi study) and thus would not have entered the rating phase (rounds 4 to 6 of the Delphi study).

From Fig. 1, it follows that most considerations are considered relevant or strongly relevant by a great majority of the panel, which confirms the need to explicitly consider process data in a wide spectrum of data governance aspects. Two items were even considered strongly relevant by more than two-thirds of the panel: *identify process-related information needs* (C.35) and *make understandable and reliable insights from process data broadly available in the organisation* (C.23). Both of these items have a strong focus on the end user, to ensure that the specific information needed to make business decisions is available and, once insights have been gathered, to ensure that end-users receive reliable analysis outcomes in an understandable way.

5.2. Implications

The insights provided in this paper have implications for the process mining domain from both an academic and practitioner viewpoints. In *academia*, the need for better understanding and improved support for data governance of process data is being increasingly recognised. For instance, process data governance recently emerged as one of the new capability areas in the broader BPM field in a Delphi study by Kerpedzhiev et al. [77], updating the original BPM capability framework of de Bruin and Rosemann [78]. Similarly, Martin et al. [55] reported several persistent challenges for the organisational use of process mining, such as poor data quality and data access barriers. Martin et al. [55] noted that giving explicit consideration to process mining in organisational data governance policies could contribute to tackling these challenges. Until now, research contributions targeting data governance of process data have focused on very specific aspects and have tended toward a strong algorithmic emphasis. Moreover, contributions have been largely clustered on the topics of process data quality and process data privacy, which cover only a small part of the data governance spectrum.

The development and further investigation of data governance aspects in the realm of process data have been recently proposed as directions for future research to ensure that high-quality data for solving business problems is always available [79]. However, there is currently a lack of understanding about what data governance for process data actually entails [16]. Such a conceptual foundation is needed to move forward and provide more comprehensive support for process data governance [80,81]. This paper marks a first step in establishing these conceptual foundations by providing an overview of data governance considerations for process data from various data governance dimensions, reflecting the views of both academics and practitioners, and specifying the key requirements for data governance in process mining. The significance of requirements in specific contexts has been highlighted as an important area of investigation in the information systems field [82], underscoring the relevance of this work. The considerations coincide with Type 1 theory according to Gregor [83] as it constitutes a starting point to further analyse and understand a novel research area. In this way, this paper significantly moves beyond the recent ImperoPD framework [16], which targets only the data quality dimension and is entirely literature-based.

The broad list of considerations opens up a plethora of research opportunities. Although selected topics, such as data quality in process mining (e.g., Martin [17]; Suriadi et al. [18]) and privacy-preserving process mining (e.g., Elkoumy et al. [19]; Fahrenkrog-Petersen et al. [84]), already received some research attention, many other data governance topics remain unexplored from a research perspective. Specific research challenges relate, for instance, to the development of dedicated methods to support the creation of a process-centric data architecture (C.1), determine critical process data (C.19), and identify process-related information needs (C.35).

From a *practitioner's* perspective, this paper clearly highlights the critical importance of considering process data in data governance strategies and policies. Moreover, the list of considerations can provide valuable guidance on how to transform process data into a strategic asset to derive the maximum benefit from the potential of process mining. In particular, the considerations could serve as an initial checklist to avoid overlooking key areas of attention. These considerations also can contribute to the design of information systems at an organisational level such that reliable process data is collected. Furthermore, this paper can provide the basis for future research focused on developing a structured roadmap for practitioners to work toward systematic process data governance.

5.3. Limitations

The results of this work need to be considered against the study's limitations. These limitations are linked to the nature of Delphi studies and the design choices made. First, the results are based on the perspectives of a limited number of experts, which is typical of Delphi studies [56]. Thus, we cannot formally exclude any form of bias in the results and cannot claim generalisability of the results. Nonetheless, the experts were carefully selected using well-defined selection criteria to ensure their expertise. Furthermore, we received positive feedback regarding the coding performed in the study as well as regarding the study itself, which supports our confidence in the results.

Second, although the ratings provide preliminary insight into the comparative relevance of data governance considerations for process data, our study does not shed light on the reasoning behind the assigned ratings. This is consistent with the exploratory character of Delphi studies [55]. Understanding the rationale behind the ratings would also require experts to give an explicit reason for each rating in the different rating rounds, which was deemed infeasible given the substantial commitment that was already expected from them.

Finally, the DAMA dimensions were provided to the panel as background information in the first round, which might have constrained the experts from providing their first input for the study. Nonetheless, the main reason for providing the DAMA dimensions at the start was to develop a shared understanding of data governance and thereby stimulate broad thinking. It constituted only background information, which was followed by an open question in which experts were asked to suggest data governance considerations for process data based on their own experience and understanding. The dimensions were not included in the questions of the first round. According to Paré et al. [23], brainstorming is unstructured and asks respondents to respond to one or more broad questions. In our first round, a single broad question was asked, as experts were requested to list data governance considerations they perceived as "either specific to process data or have a specific interpretation for process data, impacting the value that an organisation can draw from process mining." Because experts were not asked to provide input according to the DAMA dimensions, we strongly believe that providing them was not restrictive. This is also demonstrated by the fact that an additional dimension was added when coding the input from the first round, indicating that experts reflected even more broadly than the DAMA dimensions. Furthermore, qualitative feedback was received from participants indicating that DAMA dimensions helped the participants better understand what the concept of data governance entails.

6. Conclusion

Data is being increasingly considered a strategic asset by organisations, which in turn is increasing the significance of data governance. Despite the importance of data governance, it remains an under-researched area. Process data retrieved from information systems reflect the execution of organisational processes, which can be used to uncover the behaviour and performance of business operations in an organisation. The reliability of process data is critical for strategic decision making. Previous research took a process data perspective on specific data governance dimensions, such as data quality, privacy and event log construction; however, a holistic set of data governance considerations for process data was not identified. We address this gap by presenting a data governance framework for process data with 38 key considerations for 11 distinct dimensions. This framework is based on the outcomes of a Delphi study with a panel of experts from academia and industry with expertise in process mining and data governance.

To the best of our knowledge, this is the first framework that provides data governance considerations for process data, which concurs with type 1 theory according to Gregor [83]. The framework provides a solid foundation for academics and practitioners and an entry point for understanding the key considerations governing process data. It can be used to guide practitioners in developing a corporate-wide program for process data governance. The framework also provides various avenues for future research. Future research can apply this framework and validate the dimensions and considerations by studying a range of real-life process mining projects. Furthermore, the framework can be used to develop a methodology for implementing process data governance. The framework can also provide a starting point for building a process data governance maturity model. Finally, the framework identifies various areas that require attention and research to support organisations in turning process data into a strategic asset.

CRediT authorship contribution statement

Kanika Goel: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Niels Martin:** Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arthur ter Hofstede:** Writing – review & editing, Validation, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Acknowledgments

We are grateful to experts for providing rich insights in the Delphi study, which assisted in the development of the data governance framework.

Appendix A. Demographic information about the Delphi panel

This section provides a demographic overview of the experts involved in this study.
Table A.4

Table A.4
Demographic overview of experts.

Academics	11	Practitioners	10
Country of main activity			
Australia	2	Australia	3
Austria	1	Austria	–
Belgium	1	Belgium	–
Brazil	1	Brazil	–
Germany	2	Germany	1
Israel	1	Israel	–
Netherlands	1	Netherlands	5
South Korea	1	South Korea	–
Japan	–	Japan	1
Years of process mining experience			
< 5	1	< 5	5
5–10	5	5–10	1
> 10	5	> 10	4
Years of data governance experience			
< 5	8	< 5	7
5–10	2	5–10	1
> 10	1	> 10	2
Years with PhD		Years of work experience	
< 5	1	< 5	2

(continued on next page)

Table A.4 (*continued*)

Academics	11	Practitioners	10
5–10	1	5–10	2
> 5	9	> 10	6
PhD discipline		Job position	
Accounting Information Systems		Process Mining Champion	
Business Informatics		Data Governance Manager	
Computer and Systems Engineering		Data Scientist	
Computer Science		Director	
Industrial Engineering		Entreprenur	
Information Systems Engineering		Process Mining Consultant	
Information Technology		Process-Aware Data Scientist	
Process Mining		—	
Social and Economic Sciences		—	

Appendix B. Delphi study procedure

Round 1: Brainstorming of Data Governance Considerations

In round 1, conducted between August 23, 2021, and September 16, 2021, we provided the experts with the motivation for this study along with the design choices that we made. Moreover, we provided the experts background information on process mining and data governance. For data governance, we provided the definition of each of the dimensions of data governance proposed by DAMA [3]. We also requested consent from the participants to take part in the survey in accordance with the ethics agreement. We asked 26 participants to provide at least five data governance considerations for process mining. We also used the first questionnaire to obtain background information about the experts, such as their country of residence, areas of work, experience with process mining, and areas where process mining had been applied. We received 21 responses and a total of 106 data governance considerations. These 106 data governance considerations were subject to hybrid coding and consolidated into 36 considerations. The considerations were grouped according to the DAMA dimensions as applicable.

Round 2: Validation of data governance considerations

In round 2, conducted between September 24, 2021, and October 11, 2021. In this round, we provided the experts with the consolidated considerations along with their definitions and requested their feedback on our coding. We also asked for their input on any potential additions, deletions, or regrouping of consolidated considerations. Twenty-one experts participated in this round and provided their input. The experts received background information on the motivation for the study, process mining, and data governance. Based on their input, new considerations were added to dimensions, deleted from dimensions, and regrouped. This resulted in a total of 50 considerations along with the introduction of a new dimension of process data governance. In round 2, we also received initial coding and overall satisfaction ratings from the experts. We found a mean satisfaction score of 5.9 for the overall study and a mean satisfaction score of 5.62 for the second round.

Round 3: Narrowing down of data governance considerations

In round 3, conducted between October 16, 2021, and November 10, 2021, the objective was to narrow down the considerations to a manageable number. We invited all 21 participants who had provided input in round 2 to participate in round 3, of whom 20 responded. The questionnaire for this round, as before, included the motivation for the study and background information on process mining and data governance. We informed the experts that consolidation of their responses from round 2 resulted in a total of 50 considerations. We then asked the experts to select those considerations that they found relevant to construct a manageable set. We also advised them that considerations that did not receive adequate support would be removed from the study. After receiving the votes of the experts, only items deemed relevant by the majority of the panel (i.e., more than 50 % of the panel) were retained for the next round [23]. Application of this rule resulted in 38 considerations at the end of round 3. Furthermore, we obtained a mean satisfaction score of 5.95 for the overall study and a mean satisfaction score of 5.65 for the coding we did in the third round.

Round 4: Rating of data governance considerations

For round 4, conducted between November 15, 2021, and November 28, 2021, all 20 participants were invited and were asked to rate the data governance considerations to provide an overview of their comparative relevance. As in previous rounds, the questionnaire for this round included motivation for the study, background information on process mining and data governance, as well as information on this round. We provided data governance considerations and definitions and requested the experts to rate them on a 5-point Likert scale with the following options: "irrelevant," "slightly relevant," "moderately relevant," "relevant," and "strongly relevant." We included the irrelevant option to give experts a chance to still rate a consideration as "irrelevant" in case their opinion from the previous round was not taken into account. We also informed them that we would not be providing the number of votes from the previous round, to avoid bias in this and the subsequent rounds of the study. We received 18 responses out of 20 in this round. Interestingly, none of the considerations was rated "irrelevant" at the end of round 4. We found a mean satisfaction score of 6.11 for the overall study and a mean satisfaction score of 6.17 for the fourth round.

Round 5: Rating of data governance considerations

In round 5, conducted between November 30, 2021, and December 14, 2021, all 18 participants were invited. In this round, the experts were asked to rate the considerations again as in the previous round; however, in this round they were provided with the aggregate distribution of votes for each consideration from the previous round. We provided this distribution to give the experts an overview of how other experts rated the various considerations and also provide an opportunity to change the rating of their relevance. We received 17 responses at the end of this round and did not notice any drop in data governance considerations, with a mean satisfaction score of 5.94 for the overall study in this round.

Round 6: Rating of data governance considerations

Round 6, the last round, was conducted between December 15, 2021, and January 24, 2022. This round took longer than usual because of the holiday period (i.e., Christmas and New Years Day) that it spanned. All 17 participants from round 6 took part in this round. At the end of this round, as in the previous round, we did not see any change in the number of data governance considerations. The mean satisfaction score for the overall study was 5.76. Because so little change in the final results was observed over rounds 4 to 6, we decided to terminate the study with the established termination condition [23]. The convergence of the results also reinforced our decision to terminate the study at that point.

Appendix C. Intermediate coding results for the Delphi study

Appendix C.1. Round 1, 2, and 3 results

In the results of rounds 1, 2, and 3, if a consideration was included in a particular round, then the value associated with that consideration in that round was "yes." For example, process-centric data architecture was unanimously identified as a data governance consideration for process mining in all three rounds.

Table C.5

Table C.5
Round 1, 2, and 3 results.

Consideration	Definition	Round 1	Round 2	Round 3
Data Architecture				
Process-centric data architecture	There is a need for a well- documented data architecture focused, considering process mining from the start rather than an afterthought.	Yes	Yes	Yes
Data Modeling and Design				
Modeling process mining requirements while designing the data model	There is a need for the data model to capture all fields required for process mining in a format that enables the flexible use of process mining tools.	Yes	Yes	Yes
Identify and implement appropriate integrity constraints for process data	There is a need for the data model to include the necessary integrity constraints to collect appropriate process data.	Yes	Yes	Yes
Proper understanding of the data model	There is a need for in-depth and up-to-date understanding of the data model to build an event log.	Yes	Yes	Yes
Support for multiple levels of abstraction	There is a need for the data model to support multiple levels of abstraction for process data.	No	Yes	Yes
Data Storage and Operations				
Availability of redo logs	There is a need to maintain redo logs, which contain information on the historical states of data values.	Yes	Yes	No
Data storage guidelines	There is a need to maintain data storage guidelines providing information on what process data to retain and for how long, also considering its implications at the level of information systems.	Yes	Yes	Yes
Widely adopted cross-organisational event log storage standard	There is a need for a widely adopted cross-organisational event log storage standard that enables unified data and meta-data storage and also facilitates data sharing.	Yes	Yes	Yes
Data Security				
Privacy preservation versus analysis preservation	There is a need to weigh the implementation of privacy- preserving techniques for human process participants against their impact on the process mining potential.	Yes	Yes	Yes
Privacy preservation of organisation's sensitive data in a cross-organisational context	There is a need to ensure the privacy of an organisation's sensitive data when performing process mining in a cross-organisational context.	Yes	Yes	Yes
Privacy preservation of human process participants	There is a need to implement mechanisms that safeguard the privacy of human process participants whose data is analysed.	Yes	No	No
Store process data in a secure environment	There is a need to store process data in a physically safe environment with appropriate cybersecurity controls.	Yes	Yes	Yes
Data tampering policies	There is a need to create and maintain data tampering policies to be able to track illegitimate changes and loss of process data.	No	Yes	No
Risk mitigation strategies	There is a need to create and maintain risk mitigation strategies enabling handling of process data when exposed to risks (e.g., accidental loss of sensitive data or violation of privacy regulations).	No	Yes	No
Data integration and interoperability				
Customisable event log construction	There is a need to support the customisable creation of event logs which include the relevant data attributes for a specific process mining analysis.	Yes	Yes	No
Cross-system data integration	There is a need to integrate data from more than one system or database to create proper event logs, which may require data conversion, mapping, and use of knowledge graph methods.	Yes	Yes	Yes
Collection of end-to-end process data	There is a need to collect data for the end-to-end process instead of considering a fragment of the process.	Yes	Yes	Yes
Event log construction guidelines and methods	There is a need for guidelines and methods regarding event log construction, covering both technical (e.g., the extraction of process data) and non-technical (e.g., the involvement of both data and process experts) aspects.	Yes	Yes	Yes
Event log transformation guidelines and methods	There is a need for guidelines and methods supporting repeatable event log transformation to contribute to the transparency and explainability of process mining analyses.	Yes	Yes	Yes
Permanent availability of usable process data	There is a need for a permanent availability of cleansed and transformed process data.	Yes	Yes	No
Process data interoperability among organisational entities	There is a need to ensure process data interoperability between departments or partnering organisations to conduct process mining within that scope.	Yes	Yes	Yes
Documents and Content				
Keep track of documents containing relevant process data external to the database	There is a need to keep track of documents containing relevant process data external to the database in order to integrate it systematically with automatically recorded process data.	Yes	Yes	No
Ability to capture process data from documents	There is a need for techniques and tools to capture relevant process data from (potentially paper-based) documents external to the database, often comprising unstructured text.	No	Yes	No
Store a rich set of process-, user-and domain-related artefacts in a process repository	There is a need to use process data to store a rich set of process-, user-, and domain-related artefacts such as process documentation, process models, user profiles, and ontologies in a process repository.	No	Yes	Yes
Ability to capture data from multidimensional databases	There is a need to capture data from modern multidimensional databases such as document databases.	Yes	No	No
Reference and Master Data				
Master data specification	There is a need to identify and implement appropriate methods to specify master	No	Yes	Yes

(continued on next page)

Table C.5 (*continued*)

Consideration	Definition	Round 1	Round 2	Round 3
Reference data specification	data for processes. There is a need to identify and implement appropriate methods to specify reference data (e.g., range of values for an attribute) for processes as well as to connect reference data to existing domain ontologies and knowledge bases.	Yes	Yes	No
Up-to-date master and reference data	There is a need to ensure that the master and reference data is up-to-date at all times.	No	Yes	Yes
Identify critical process data	There is a need to identify critical master and reference data (e.g., set of activities, events, or values of attributes) to get maximal value from process data governance efforts.	Yes	Yes	Yes
Data Warehousing and Business Intelligence				
Data warehouse for process mining	There is a need to maintain a data warehouse for process mining, potentially integrated in a data lake, ensuring that process data from various systems is available in a schema that facilitates usage of process mining tools.	Yes	Yes	Yes
Availability of real-time process data Build support for	There is a need to ensure the availability of real-time process data to support decision making.	No	Yes	Yes
process mining	There is a need to make dashboards and other (preferably multipurpose) tools available to build organisational support for collecting, storing, and analysing process data.	Yes	Yes	Yes
Make insights from process data insights broadly available in the organisation	There is a need to make understandable and reliable process data insights available to knowledge workers and decision makers to enable monitoring and analysis of processes as well as for decision making.	Yes	Yes	Yes
Maintain a repository of process mining artefacts	There is a need to maintain a repository of process mining artefacts, which can be leveraged in future analyses.	No	Yes	Yes
Meta-data				
Repository of process-related meta-data	There is a need to establish and maintain a complete, up-to-date, and easily accessible repository of process-related meta-data, which provides detailed information regarding the meaning of process data (e.g., a glossary providing the interpretation of all attributes and their values).	Yes	Yes	Yes
Document the mapping between real-life process concepts and data recording	There is a need for careful documentation and a perfect understanding of all domain expertise regarding how real-life process concepts such as tasks are captured as process data. There is a need to carefully document data provenance to make the source and transformation of process data fully traceable.	Yes	Yes	Yes
Document process data provenance	There is a need to carefully document the internal controls that are in place related to process data.	No	Yes	Yes
Document internal controls related to process data		No	Yes	Yes
Data Quality				
Verify process Data integrity automatically	There is a need to implement mechanisms that automatically verify integrity constraints related to process data.	Yes	Yes	No
Awareness of specific data quality issues for process data	There is a need for awareness for data quality issues that are specific to process data to allow the use of reliable and unambiguous data.	Yes	Yes	Yes
Data quality assessment guidelines and rules for process data	There is a need to maintain and document guidelines and rules regarding the assessment of data quality issues in process data, taking into account the use case at hand.	Yes	Yes	Yes
Assure the quality of process data transformations	There is a need to assure the quality of process data transformations, given their great impact on process mining outcomes.	Yes	Yes	Yes
Build trust in the truthfulness of process data	There is a need to build trust in the truthfulness of process data to gain the confidence of business users in process mining results.	Yes	Yes	Yes
Use process mining when identifying/handling quality issues in process data	There is a need to consider the use of process mining algorithms when identifying/handling quality issues in process data as it enables the generation of novel insights in process data quality.	Yes	Yes	No
Supporting Organisational Policies and Programs				
Ensure the consistency of process data governance policies and programs	There is a need to ensure the consistency of process data governance policies and programs across different dimensions due to their close interconnection.	No	Yes	Yes
Training and communication programs for human resources	There is a need to develop training and communication programs for human resources to correctly use systems, enabling the collection of reliable process data.	Yes	Yes	Yes
Identify process-related information needs	There is a need to identify the information needed to make business decisions regarding processes and align process data	Yes	Yes	Yes
Agile process data governance	governance efforts with it. There is a need for stepwise, agile, and lean introduction of process data governance considerations such that the organisation can witness the value of process data.	Yes	Yes	No
Process data and meta-data ownership	There is a need to assign dedicated responsibilities within the organisation regarding process data and meta-data.	Yes	Yes	Yes
Explicitly link strategic priorities to process data	There is a need to explicitly link the organisation's strategic priorities to the process data that can be leveraged to support achieving them.	No	Yes	Yes
Centre of excellence for process data analysis	There is a need to facilitate the development of a centre of excellence on process data analysis within the organisation.	No	Yes	Yes
Identify and apply best practices	There is a need to identify and apply best practices to devise and implement data governance considerations for process data.	No	Yes	Yes

Appendix C.2. Round 4, 5, and 6 results

The total number of considerations remained the same across rounds 4, 5, and 6. This section displays the median (*median*) and mode (*mode*) rating per consideration for the three rounds.

Table C.6

Table C.6
Round 4, 5, and 6 scores.

Data storage guidelines	R	R	R	R	R	R
Widely adopted cross-organizational event log storage standard	R	R	R	R	R	R
Privacy preservation versus analysis preservation	R	R	R	R	R	R
Privacy preservation of organisation's sensitive data in a cross-organisational context	R	R	R	STR	R	STR
Store process data in a secure environment	R, STR	STR	STR	STR	STR	STR
Cross-system data integration	R	R	R	R	R	R
Collection of end-to-end process data	R	STR	R	MR	R	STR
Event log construction guidelines and methods	R	R	R	R	R	R
Event log transformation guidelines and methods	R	MR	R	MR	R	MR
Process data interoperability between organisational entities	R	R	R	R	R	R
Store a rich set of process-, user-, and domain-related artefacts in a process repository	MR	MR	MR	MR	MR	MR
Master data specification	R	R	R	R	R	R
Up-to-date master and reference data	R, STR	STRR	STR R	STR R	STR	STR
Identify critical process data	R				R	R
Data warehouse for process mining	R	STR	R	STR	R	R
Availability of real-time process data	MR, R	MR	MR	MR	MR	MR
Build support for process mining	R	R	R	R	R	R
Make insights from process data broadly available in the organisation	R, STR	STR	STR	STR	STR	STR
Maintain a repository of process mining artefacts	R	R	R	R	R	R
Repository of process-related metadata	R	R	R	R	R	R
Document the mapping between real-life process concepts and data recording	R	R	R	R	R	R
Document process data provenance	R	R	R	R	R	R
Document internal controls related to process data	R	R	R	R	R	R
Awareness of specific data quality issues for process data	R	R	R	R	R	R
Data quality assessment guidelines and rules for process data	R	R	R	R	R	R
Assure the quality of process data transformations	R	R	R	R	R	R
Build trust in the truthfulness of process data	R	R	R	R	R	R
Ensure the consistency of process data governance policies and programs	R	STR	R	R	R	R
Training and communication programs for human resources	R	R	R	R	R	R
Identify process-related information needs	R, STR	STR	STR	STR	STR	STR
Process data and metadata ownership	R	R	R	R	R	R
Explicitly link strategic priorities to process data	R	R	R	R	R	R
Centre of excellence for process data analysis	MR	MR	MR	MR	MR	MR

IRR, irrelevant; SLR, slightly relevant; MR, moderately relevant; R, relevant; STR, strongly relevant.

References

[1] R. Abraham, J. Schneider, J. Vom Brocke, Data governance: a con- ceptual framework, structured review, and research agenda, Int. J. Inf. Manage 49 (2019) 424–438.
[2] J. Ladley, Data governance: How to design, deploy, and Sustain an Effective Data Governance Program, Academic Press, 2019.
[3] M. Mosley, M.H. Brackett, S. Earley, D. Henderson, The DAMA guide to the data management body of knowledge. Technics Publications, 2010.
[4] A.G. Carretero, F. Gualo, I. Caballero, M. Piattini, MAMD 2.0: environment for data quality processes implantation based on ISO 8000- 6X and ISO/IEC 33000, Comput. Stand. Interfaces. 54 (2017) 139–151.
[5] B. Otto, Organizing data governance: findings from the telecommu- nications industry and consequences for large service providers, Commun. Assoc. Inf. Syst. 29 (2011) 3.
[6] F. Haneem, N. Kama, N. Taskin, D. Pauleen, N.A.A. Bakar, De- terminants of master data management adoption by local government organizations: an empirical study, Int. J. Inf. Manage 45 (2019) 25–43.
[7] Statista, 2019. Digital economy compass 2019. https://www.statista.com/study/52194/digital-economy-compass/#0.
[8] M. Al-Ruithe, E. Benkhelifa, K. Hameed, A systematic literature review of data governance and cloud data governance, Person. Ubiquit. Comput. 23 (2019) 839–859.
[9] V. Holt, M. Ramage, K. Kear, N. Heap, The usage of best practices and procedures in the database community, Inf. Syst. 49 (2015) 163–181.
[10] C. Begg, T. Caira, Exploring the SME quandary: data governance in practise in the small to medium-sized enterprise sector, Electron. J. Inf. Syst. Eval. 15 (2012) 3–13.
[11] Q. Zhang, X. Sun, M. Zhang, Data matters: a strategic action framework for data governance, Inf. Manag. 59 (2022) 103642.
[12] W.M.P. van der Aalst, Process Mining: Data Science in Action, Springer, 2016.
[13] M. Dumas, W.M.P. van der Aalst, A.H.M. ter Hofstede (Eds.), Process-Aware Information Systems: Bridging People and Software through Process Technology, Wiley, 2005, https://doi.org/10.1002/0471741442. URL.
[14] C. dos Santos Garcia, A. Meincheim, E.R.F. Junior, M.R. Dallagassa, D.M.V. Sato, D. R. Carvalho, E.A.P. Santos, E.E. Scalabrin, Pro- cess mining techniques and applications–a systematic mapping study, Expert Syst. Appl. 133 (2019) 260–295.
[15] M. Kerremans, Process Mining Can Support Business Operations in Driving Resilient Growth, Gartner, 2022.
[16] K. Goel, F. Emamjome, A.H.M. ter Hofstede, Data governance for managing data quality in process mining, in: J.S. Valacich, A. Barua, R. Wright, A. Kankanhalli, X. Li, S. Miranda (Eds.), Proceedings of the 42nd International Conference on Information Systems, ICIS 2021, Build- ing Sustainability and Resilience with IS: A Call for Action, Austin, TX, USA, Association for Information Systems, 2021. December 12–15, 2021URL: https://aisel.aisnet.org/icis2021/governance/governance/9.
[17] N. Martin, Data quality in process mining, in: C. Fernandez-Llatas (Ed.), Interactive Process Mining in Healthcare, Springer, 2021, pp. 53–79.
[18] S. Suriadi, R. Andrews, A.H.M. ter Hofstede, M.T. Wynn, Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs, Inf. Syst. 64 (2017) 132–150.
[19] G. Elkoumy, S.A. Fahrenkrog-Petersen, M.F. Sani, A. Koschmider, F. Mannhardt, S. N. Von Voigt, M. Rafiei, L.V. Waldthausen, Privacy and confidentiality in process mining: threats and research challenges, ACM Trans. Manag. Inf. Syst. (TMIS) 13 (2021) 1–17.
[20] A. Pika, M.T. Wynn, S. Budiono, A.H. ter Hofstede, W.M.P. van der Aalst, H. A. Reijers, Privacy-preserving process mining in health- care, Int. J. Environ. Res. Public Health 17 (2020) 1612.
[21] R. van Cruchten, H. Weigand, Towards event log management for process mining-vision and research challenges, in: International Conference on Research Challenges in Information Science, Springer, 2022, pp. 197–213.
[22] M. Jans, P. Soffer, T. Jouck, Building a valuable event log for pro- cess mining: an experimental exploration of a guided process, Enterp. Inf. Syst. 13 (2019) 601–630.
[23] G. Paré, A.F. Cameron, P. Poba-Nzaou, M. Templier, A systematic assessment of rigor in information systems ranking-type Delphi studies, Inf. Manag. 50 (2013) 207–217.
[24] P. Brous, M. Janssen, R. Vilminko-Heikkinen, Coordinating decision- making in data management activities: a systematic review of data gov- ernance principles, in: Electronic Government: 15th IFIP WG 8.5 Inter- national Conference, EGOV 2016, Guimarães, Portugal, September 5-8, 2016, Proceedings 15, Springer, 2016, pp. 115–125.
[25] I. Alhassan, D. Sammon, M. Daly, Data governance activities: an analysis of the literature, J. Decis. Syst. 25 (2016) 64–75.
[26] T. van den Broek, A.F. van Veenstra, Governance of big data collabo- rations: how to balance regulatory compliance and disruptive innovation, Technol. Forecast. Soc. Change 129 (2018) 330–338.
[27] D. Sjödin, V. Parida, M. Kohtamäki, J. Wincent, An agile co- creation process for digital servitization: a micro-service innovation ap- proach, J. Bus. Res. 112 (2020) 478–491.
[28] H.W. Volberda, S. Khanagha, C. Baden-Fuller, O.R. Mihalache, J. Birkin- shaw, Strategizing in a digital world: overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms, Long. Range Plann. 54 (2021) 102110.
[29] A. Al-Badi, A. Tarhini, A.I. Khan, Exploring big data governance frameworks, Procedia Comput. Sci. 141 (2018) 271–277.

[30] M. Al-Ruithe, E. Benkhelifa, K. Hameed, A conceptual framework for designing data governance for cloud computing, Procedia Comput. Sci. 94 (2016) 160–167.
[31] DAMA International, 2022. About us. URL: https://www.dama.org/cpages/mission-vision-purpose-and-goals.
[32] T.H. Davenport, J.E. Short, et al., Technical Report, Center for Information Systems Research, Sloan School of Management, 1990.
[33] W.M.P. van der Aalst, Process mining: a 360 degree overview, in: W.M.P. van der Aalst, J. Carmona (Eds.), Process Mining Handbook, Springer, 2022, pp. 3–34, volume 448 of *Lecture Notes in Business Information Processing*.
[34] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, F.M. Maggi, A. Marrella, M. Mecella, A. Soo, Automated discovery of process models from event logs: review and benchmark, IEEe Trans. Knowl. Data Eng. 31 (2018) 686–705.
[35] W.M.P. van der Aalst, H.A. Reijers, M. Song, Discovering social net- works from event logs, Comput. Supp. Cooperat. Work (CSCW) 14 (2005) 549–593.
[36] M. Song, W.M.P. van der Aalst, Towards comprehensive support for organizational mining, Decis. Support. Syst. 46 (2008) 300–317.
[37] J. Carmona, B. van Dongen, A. Solti, M. Weidlich, Conformance Checking, Springer, Switzerland, 2018 [Google Scholar].
[38] S. Suriadi, C. Ouyang, W.M. Van Der Aalst, A.H. Ter Hofstede, Event interval analysis: why do processes take time? Decis. Support. Syst. 79 (2015) 77–98.
[39] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, J. Karnon, Process mining for clinical processes: a comparative analysis of four australian hospitals, ACM. Trans. Manage Inf. Syst. 5 (2015) 1–18.
[40] M.T. Wynn, E. Poppe, J. Xu, A.H.M. ter Hofstede, R. Brown, A. Pini, W.M.P. van der Aalst, ProcessProfiler3D: a visualisation framework for log-based process performance comparison, Decis. Support. Syst. 100 (2017) 93–108.
[41] C. Di Francescomarino, C. Ghidini, Predictive process monitoring, in: Process Mining Handbook, 448, LNBIP, 2022, pp. 320–346.
[42] G. Park, D. Schuster, W.M. van der Aalst, Pattern-based action engine: generating process management actions using temporal patterns of process-centric problems, Comput. Ind. 153 (2023) 104020.
[43] K. Goel, S.J. Leemans, N. Martin, M.T. Wynn, Quality-informed process mining: a case for standardised data quality annotations, ACM Trans. Knowl. Discov. Data (TKDD) 16 (2022) 1–47.
[44] M.T. Wynn, S. Sadiq, Responsible process mining-a data quality per- spective, in: International Conference on Business Process Management, Springer, 2019, pp. 10–15.
[45] M. Rafiei, W.M.P. van der Aalst, Privacy-preserving data publish- ing in process mining, in: International Conference on Business Process Management, Springer, 2020, pp. 122–138.
[46] E.G.L. de Murillas, H.A. Reijers, W.M.P. van der Aalst, Connecting databases with process mining: a meta model and toolset, in: R. Schmidt, W. Gućedria, I. Bider, S. Guerreiro (Eds.), Enterprise, Business-Process and Information Systems Modeling - 17th International Conference, BP- MDS 2016, 21st International Conference, EMMSAD 2016, Held at CAiSE 2016, Ljubljana, Slovenia, June 13-14, 2016, Proceedings, Springer, 2016, pp. 231–249, https://doi.org/10.1007/978-3-319-39429-9. URL\15.
[47] S. Zuboff, In the Age of the Smart Machine: The Future of Work and Power, Basic Books, Inc, 1988.
[48] A. Burton-Jones, What have we learned from the smart machine? Inf. Org. 24 (2014) 71–105.
[49] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, Manage Sci. 9 (1963) 458–467.
[50] U.G. Gupta, R.E. Clarke, Theory and applications of the Delphi technique: a bibliography (1975–1994), Technol. Forecast. Soc. Change 53 (1996) 185–211.
[51] S. Keeney, F. Hasson, H. McKenna, Consulting the oracle: ten lessons from using the Delphi technique in nursing research, J. Adv. Nurs. 53 (2006) 205–212.
[52] T. Schmiedel, J. vom Brocke, J. Recker, Which cultural values matter to business process management?: results from a global Delphi study, Bus. Process. Manag. J. 19 (2013) 292–317, https://doi.org/10.1108/14637151311308321. URL:.
[53] U.M. Konig, A. Linhart, M. Röglinger, Why do business processes deviate? Results from a Delphi study, Bus. Res. 12 (2019) 425–453.
[54] G. Hübner-Blöder, E. Ammenwerth, Key performance indicators to benchmark hospital information systems–a Delphi study, Methods Inf. Med. 48 (2009) 508–518.
[55] N. Martin, D.A. Fischer, G.D. Kerpedzhiev, K. Goel, S.J. Leemans, M. Röglinger, W. M.P. van der Aalst, M. Dumas, M. La Rosa, M.T. Wynn, Opportunities and challenges for process mining in organizations: results of a Delphi study, Bus. Inf. Syst. Eng. 63 (2021) 511–527.
[56] C. Okoli, S.D. Pawlowski, The Delphi method as a research tool: an example, design considerations and applications, Inf. Manag. 42 (2004) 15–29.
[57] R.C. Schmidt, Managing Delphi surveys using nonparametric statis- tical techniques, Decis. Sci. 28 (1997) 763–774.
[58] J. Kobus, M. Westner, Ranking-type delphi studies in is research: step-by-step guide and analytical extension, in: IADIS International Conference on Information System, 2016, pp. 28–38.
[59] J.F. Yeung, A.P. Chan, D.W. Chan, Developing a performance index for relationship-based construction projects in Australia: Delphi study, J. Manag. Eng. 25 (2009) 59–68.
[60] A.C. Skinner, E.M. Perrin, L.A. Moss, J.A. Skelton, Car- diometabolic risks and severity of obesity in children and young adults, N. Engl. J. Med. 373 (2015) 1307–1317.
[61] J. Saldaña, The Coding Manual for Qualitative Researchers, Sage, 2021.
[62] J. Fereday, E. Muir-Cochrane, Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development, Int. J. Qual. Methods 5 (2006) 80–92.
[63] R.E. Boyatzis, Transforming Qualitative Information: Thematic Analysis and Code Development, SAGE, 1998.
[64] J.M. Corbin, A. Strauss, Grounded theory research: procedures, canons, and evaluative criteria, Qual. Sociol. 13 (1990) 3–21.
[65] M.J. Bardecki, Participants' response to the Delphi method: an attitudinal perspective, Technol. Forecast. Soc. Change 25 (1984) 281–292.
[66] H.A. Linstone, M. Turoff (Eds.), The Delphi Method: Techniques and Applications, Addison-Wesley, Reading, MA, 1975.
[67] W.M.P. van der Aalst, A. Adriansyah, A.K.A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, R.P.J.C. Bose, P. van den Brand, R. Brandtjen, J.C.A.M. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, J.E. Cook, N. Costantini, F. Curbera, E. Damiani, M. de Leoni, P. Delias, B.F. van Dongen, M. Dumas, S. Dustdar, D. Fahland, D.R. Ferreira, W. Gaaloul, F. van Geffen, S. Goel, C. W. Günther, A. Guzzo, P. Harmon, A.H.M. ter Hofstede, J. Hoogland, J. E. Ingvaldsen, K. Kato, R. Kuhn, A. Kumar, M.L. Rosa, F.M. Maggi, D. Malerba, R. S. Mans, A. Manuel, M. McCreesh, P. Mello, J. Mendling, M. Montali, H.R. M. Nezhad, M. zur Muehlen, J. Munoz-Gama, L. Pontieri, J. Ribeiro, A. Rozinat, H. S. Pérez, R.S. Pérez, M. Sepulveda, J. Sinur, P. Soffer, M. Song, A. Sperduti, G. Stilo, C. Stoel, K.D. Swenson, M. Talamo, W. Tan, C. Turner, J. Vanthienen, G. Varvaressos, E. Verbeek, M. Verdonk, R. Vigo, J. Wang, B. Weber, M. Weidlich, T. Weijters, L. Wen, M. Westergaard, M.T. Wynn, Process mining manifesto, Lect. Notes Bus. Inf. Process. 99 (2011) 169–194.
[68] G. Schuh, A. Gützlaff, S. Cremer, S. Schmitz, A. Ayati, A data model to apply process mining in end-to-end order processing processes of manufacturing companies, in: 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), IEEE, 2020, pp. 151–155.
[69] D.A. Fischer, K. Goel, R. Andrews, C. van Dun, M.T. Wynn, M. Röglinger, Towards interactive event log forensics: detecting and quanti- fying timestamp imperfections, Inf. Syst. 109 (2022) 102039.
[70] R. Andrews, F. Emamjome, A.H. ter Hofstede, H.A. Reijers, Root- cause analysis of process-data quality problems, J. Bus. Anal. 5 (2022) 51–75.
[71] T. Vogelgesang, H.J. Appelrath, Pmcube: a data-warehouse- based approach for multidimensional process mining, in: Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, Springer, 2016, pp. 167–178. August 31–September 3, 2015, Revised Papers 13.
[72] G. Ibanez-Sanchez, C. Fernandez-Llatas, Z. Valero-Ramon, J.L. Bayo-Monton, Pmapp: an interactive process mining toolkit for building healthcare dashboards, in: International Workshop on Explainable Artificial Intelligence in Healthcare, Springer, 2023, pp. 75–86.
[73] V. Nqampoyi, L.F. Seymour, D.S. Laar, Effective business process management centres of excellence, in: Research and Practical Issues of Enterprise Information Systems: 10th IFIP WG 8.9 Working Conference, CONFENIS 2016, Vienna, Austria, December 13–14, 2016, Proceedings 10, Springer, 2016, pp. 207–222.
[74] Reinkemeyer, L., 2020. Process mining in action. Process Mining in Action Principles, Use Cases and Outloook.
[75] R.J.C. Bose, R.S. Mans, W.M.P. van der Aalst, Wanna improve process mining results?, in: Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining IEEE, 2013, pp. 127–134.
[76] C. Liu, H. Duan, Q. Zeng, M. Zhou, F. Lu, J. Cheng, Towards comprehensive support for privacy preservation cross-organization business process mining, IEEe Trans. Serv. Comput. 12 (2016) 639–653.
[77] G.D. Kerpedzhiev, U.M. König, M. Röglinger, M. Rosemann, An exploration into future business process management capabilities in view of digitalization: results from a Delphi study, Bus. Inf. Syst. Eng. 63 (2021) 83–96.
[78] T. de Bruin, M. Rosemann, Using the Delphi technique to identify BPM capability areas, in: Proceedings of the 18th Australasian Conference on Information Systems, 2007, pp. 643–653.
[79] A.H.M. ter Hofstede, A. Koschmider, A. Marrella, R. Andrews, D.A. Fischer, S. Sadeghianasl, M.T. Wynn, M. Comuzzi, J. De Weerdt, K. Goel, et al., Process-data quality: the true frontier of process mining, ACM. J. Data Inf. Qual. (2023).
[80] J. Gerring, Social Science methodology: A unified Framework, Cam- bridge University Press, 2011.
[81] K. Goel, W. Bandara, G. Gable, Conceptualizing business process standardization: a review and synthesis, Schmalenbach J. Bus. Res. (2023) 1–43.
[82] C. Chua, M. Indulska, R. Lukyanenko, W. Maass, V.C. Storey, Misq research curation on data management, MIS Q. (2022) 1–12.
[83] S. Gregor, The nature of theory in information systems, MIS. Q. 30 (2006) 611–642. URL: http://misq.org/the-nature-of-theory-in-information-systems.html.
[84] S.A. Fahrenkrog-Petersen, H. Van Der Aa, M. Weidlich, Pretsa: event log sanitization for privacy-aware process discovery, in: Proceedings of the 2019 International Conference on Process Mining, IEEE, 2019, pp. 1–8.

Dr Kanika Goel is a lecturer at the School of Information Systems at the Queensland University of Technology. Her research interests include process management, data governance, data analytics, and process-oriented data mining. She is interested in exploring the potential of data in transforming business processes and had led several industry projects. Her research has been published in highly regarded journals such as Business & Information Systems Engineering and Information Systems. Kanika has presented her work at several national and international conferences. She Kanika is a member of IEEE Task Force on Process Mining and BPM Association.

Niels Martin, Ph.D., is an Assistant Professor at the research group Business Informatics of the Faculty of Business Economics at Hasselt University (Belgium). He is also affiliated to Research Foundation Flanders (FWO) as a postdoctoral researcher. At Hasselt University, he is currently the lead of the research cell Process Analytics in Healthcare, focusing on topics such as diagnostic analytics, data quality and capacity management. His-research has been published in highly-regarded journals such as IEEE Transactions on Services Computing, Decision Support Systems and Business & Information Systems Engineering. Moreover, he presented research results at scientific conferences such as the Winter Simulation Conference and workshops of the International Conference on Business Process Management. As an active community member, he is steering committee member of the Process-Oriented Data Science for Healthcare Alliance, and member of the IEEE Task Force on Process Mining and the Scientific Research Community on Process Mining.

Prof. Arthur ter Hofstede received his PhD degree from the Katholieke Universiteit Nijmegen (since renamed to Radboud Universiteit), Nijmegen, The Netherlands, in 1993. He has worked at Queensland University of Technology, Brisbane, Australia, since 1997, and is currently a professor and the Head of the School of Information Systems. From February 2010 to January 2018, he was also a part-time professor at the Information Systems Group of Eindhoven University of Technology in Eindhoven, The Netherlands. In 2010 and 2011 he was a Senior Visiting Scholar at Tsinghua University in Beijing, China. In May 2010 he was a visiting professor at University of Rome 'La Sapienza'. Arthur ter Hofstede's research interests lie in the area of business process management, in particular business process automation and process mining. He was involved in the well-known workflow patterns initiative (www.workflowpatterns.com) and at QUT he has managed the well-known YAWL initiative (www.yawlfoundation.org). Arthur ter Hofstede is a co-author on over 200 publications, including over 90 journal publications.