

Process mining on a public procurement dataset: a case study

Roberto Nai¹[0000–0003–4031–5376], Emilio Sulis¹[0000–0003–1746–3733], Rosa Meo¹[0000–0002–0434–4850], Francesco Gorgerino²[0000–0002–2441–4027], Gabriella Margherita Racca²[0000–0002–6527–7794], and Laura Genga³[0000–0001–8746–8826]

¹ Computer Science Department - University of Turin

{roberto.nai,emilio.sulis,rosa.meo}@unito.com

² Management Department - University of Turin

{francesco.gorgerino,gabriella.racca@unito.it}@unito.com

³ Eindhoven University of Technology (TU/e), The Netherlands

l.genga@tue.nl

Abstract. Legal informatics increasingly explores new frontiers due to the digitization of law. The availability of legal activities data in information systems enables the exploitation of computational technologies and algorithms. This is the case of mining event logs to automatically derive meaningful knowledge. The paper focuses on the automated analysis of a large dataset of public procurements, contract awards and complaint procedures. Process discovery techniques are applied on a real legal dataset of Italian procurement processes. In particular, variant analysis is explored to detect meaningful differences in the process variants. The results demonstrate the significance of applying automated log file analysis techniques in the legal field.

Keywords: Public procurement process · Process mining · Variant analysis · Legal dataset

1 Introduction

The digitization of legal activities led to an increasing availability of data, typically registered in information systems of public organizations such as public administration or judicial bodies, e.g. courts [9, 10, 29]. The abundance of data is coupled with the advent of new computational technologies, which led to the emergence of innovative solutions to support the automated analysis of temporal data. In this perspective, a relatively new approach for providing knowledge about data registered in information systems is *process mining* (PM), aimed at discovering, monitoring and analyzing organization processes exploiting data generated during the process execution, stored in the form of *event logs* [1]. As a bridge between data mining and business process management (BPM), the discipline provides a meaningful process-oriented perspective.

The areas of application are many, from industry to health care; however, there are not many such studies in the analysis of legal data. For these sectors it is

increasingly highlighted the importance of collaboration between data scientists and lawyers, starting from their education which must be based on interdisciplinarity [19]. Nevertheless, temporal series of legal events can be investigated to *automatically discover* and visualize the ordering of execution (or *control-flow*) of activities belonging to a legal process, such as the awarding of a procurement, or the issuance of a sentence. Such a model describes how a process is *actually* carried out in real-life contexts. This represents valuable knowledge within the legal domain, where processes tend to be quite complex, often appropriate and updated documentation is not available. This poses critical challenges when monitoring the process executions (for example, to verify whether they fulfil a set of requirements) or when looking for opportunities to improve the outcome or the process efficiency [11, 12, 16, 32, 35]. Another crucial aspect to consider when analyzing legal processes is their potential heterogeneity. Indeed, the executions of these processes can vary quite substantially among each other, depending on the decisions taken by the human actors or on the contextual factors.

This intrinsic heterogeneity often hampers the results obtained by applying process discovery techniques, since generating a single model from a set of heterogeneous executions can easily hide some important characteristics of specific process variants. In fact, also the legal literature considers important to analyse the timing of tender procedures by type, by threshold and with regard to the presence or absence of litigation [7, 12].

As a case study, we focused on the Italian public procurement process in the period 2016-2022. The time frame is fairly consistent with the validity of the Italian public procurement code which came into force in April 2016 and was repealed on 31 March 2023⁴. In this research, we propose an exploration of PM techniques, in particular *variant analysis*⁵ techniques, to grasp relevant insights on legal processes while dealing with their intrinsic heterogeneity. We argue that understanding the differences between variations in legal processes helps analysts to make informed decisions about how to standardize or improve a judicial process. We focus on the following research questions: RQ1) Can we obtain meaningful insights from applying process discovery techniques on legal data from information systems?; RQ2) Can we identify execution variants of a legal process showing relevant differences from the main process behaviours by means of variant analysis techniques?

Given the exploratory and application-driven nature of our research, we have chosen a real-world case study to carry out our investigation, following the principles of *case research* methodology [40]. As a case study, we apply PM techniques on a real legal dataset about procurement process in Italy. First, process discovery techniques are used to automatically obtain the control-flow of procurement notices, consisting of a dozen of timed activities from the procurement’s issuance, the participation notices, up to the execution and termination of the procedure.

⁴ Legislative Decree n. 50 of 2016, replaced by the Legislative Decree n. 36 of 2023

⁵ On this point, a linguistic disambiguation is necessary: in the article we speak of “variant” both in the PM sense (with regard to “variant analysis”), and in the legal sense of variant of the object of the contract (art. 45 of the directive 2014/24/EU)

Furthermore, we identified several features corresponding to relevant contextual factors (sector, duration, costs, presence of complaints) to determine whether they can be used to determine meaningful process variants. We then apply variant analysis techniques to further investigate differences among the identified variants.

The paper first provides an overview of related work and introduces the real-world legal scenario, in Section 2. Then, we describe our methodology with proposed process discovery and variant analysis techniques in Section 4. We report on the results and evaluation carried out in Section 5 and we finally conclude in Section 6.

2 Related work

Legal aspects have been typically addressed in a BPM perspective by considering regulatory compliance issues [18, 33]. Some works were proposed to automatically check the conformance of processes, but an effort of manual supervision remains necessary [4, 20]. Recently, the PM discipline gains attention for several application areas, from the typical BPM research area [1, 41], healthcare [24], or education [5, 25]. The conformance checking research seems promising to investigate legal compliance issues [8, 30].

To the best of our knowledge, very few works investigated a process-oriented approach to legal cases. At the intersection between PM and law, in [22] the authors introduced a framework to describe the GDPR impact for the design of PM systems, while [14] discusses approaches to ensure privacy in PM. Some works explore real-world cases of process discovery involving public procurement: a case study focuses on a heuristic algorithm revealing a concept drift in publication of contract in the Philippines [34], while [31] investigates public procurement procedures in Croatia.

A comprehensive application of process discovery in the legal field is [38], where authors applied discovery techniques for the extraction of lawsuit processes from the information system of the Court of Justice of the State of Sao Paulo, Brazil; they demonstrate how to provide insights for improvement to increase judicial productivity.

Artificial intelligence techniques have been also explored in legal domain, including machine learning techniques [26], merging dataset [27], explainability [23]. We build on these works and add an application of variant analysis techniques, whose goal consists in supporting the user in identifying and comparing different variants of process executions to assess and qualify their differences [37]. Informally, a process variant identifies a subset of process executions sharing one or multiple properties that distinguish them from other executions. The decision on which properties to select depends on the process under investigation, and it is left up to the analyst. Several approaches have been developed in literature in the last few years. Some approaches aim at supporting the comparison of process models mined from different variants [13, 21]. Others focus on the direct comparison of different event logs [39]. Finally, a third group of techniques takes

into account the event logs and the corresponding process models together to explain variant differences [6, 36]. In this work, we focus on the latter category since it usually provides richer insights. More precisely, we show an application of the approach proposed by Bolt et al. [6], where visualization techniques are combined with statistical analysis to assess the robustness of the derived insights. We demonstrate the feasibility of such analysis on a real event log of a procurement creation process in the legal field.

3 Case study

Our work exploits a legal dataset involving the public procurement process in Italy. The National Anti-Corruption Authority⁶ (ANAC) is an independent Italian administrative authority whose task is to prevent corruption in the Italian public administration, in particular in public procurement. Since 2010 ANAC manages the National Public Contracts Database (henceforth “BDNCP”)⁷. Its most important dataset includes all the calls for procurement⁸, i.e. 2,598,533, in the period 2016-2022. From this main dataset, we extracted the procurements created by the *municipalities* (a town or city that has local government), i.e. 511,733 tenders (representing 27% of the total). Another dataset from Italian Administrative Justice (IAJ)⁹ includes the judges’ sentences related to the public procurement complaints.

Dataset overview In the ANAC dataset, each procurement is identified by an alphanumeric value called CIG (the key value) and it has the following relevant features: the *sector* to which it belongs, of three different values: “services” (48%), “works” (42%), “supplies” (10%); the administrative *region* of the municipality that issued the procurement. Italy comprises 21 regions, the most populated being Lombardy (17,8%), Piedmont (8.5%), Lazio (7.2%), Veneto (7%) and Tuscany (6.5%); the *amount* of the procurement, also classified by Italian law between “under-threshold” (up to 140k euro for services and goods; up or equal to 5,382,000 euro for works) and “above-threshold” (over 140k euro for services and goods; over 5,382,000 euro for works)¹⁰.

The IAJ dataset includes the sentences linked to the procurement referred to via the shared CIG key. In fact, a small part of public procurements suffers a complaint (litigation), resulting in a sentence (217 confirmed cases).

The legal process and the activities The data available in the ANAC dataset represent a dozen of events (activities) in the process: a procurement is cre-

⁶ <https://www.anticorruzione.it>

⁷ Art. 62-bis of the Legislative Decree no. 82 of 2005, added by Art. 44 of the Legislative Decree of 30 December 2010, no. 235

⁸ <https://dati.anticorruzione.it/opensdata>

⁹ <https://www.giustizia-amministrativa.it>

¹⁰ Commission Delegated Regulation (EU) 2021/1950 of 10 November 2021

ated (CREATION event) and published on the Italian¹¹ or European (see next subsection) communication channels of the public administration (respectively, PUBLICATION-IT or PUBLICATION-EU). A procurement is then won (WIN) by an economic operator (a private company), which is followed by the start of a contract execution (CONTRACT-START). After the start of the contract various combinations of sub-events may occur: an economic operator may sub-contract the procurement (SUBCONTRACT); procurement execution may be suspended due to problems or unforeseen events (SUSPENSION and REPRISE); the event to be performed may vary from the original bid (VARIANT). Depending on the procurement, its status may be cyclically recorded (STATE) or checks related to the status of the procurement may be performed (TEST). The procurement cycle is usually concluded at the end of the execution of the contract (CONTRACT-END) or following its ‘concluded’ status. By cross-referencing ANAC data with IAJ data, it is possible to link a further status, i.e. a court judgment (RECOURSE - meaning the complaint - and then SENTENCE) following a dispute between a supplier and a public administration concerning a specific procurement (identified by the key value CIG).

Italian case vs. European case The “Tenders Electronic Daily” (TED) website¹² publishes 676 thousand procurement notices a year, including 258 thousand calls for tenders which worth approximately 670 billion euro. In this context, the events (e.g., CREATION, PUBLICATION-* or WIN) and the main features (e.g., sector, regions/NUTS¹³, or amounts) of the ANAC dataset find a generalization in the CSV schemas of the calls for tenders of other European countries. The ANAC dataset is a specialization of EU data to which national laws apply (e.g. STATE, VARIANT or TEST). The publication of tender notices on the TED portal is mandatory for “above-threshold” amounts (see above), according to EU directives¹⁴. Regarding the RECOURSE and SENTENCE states, in all countries there is the possibility of calling a review for a public procurement award and obtaining a judgment¹⁵ with a certain date (hence the event); in some states such as France and Spain, the judicial review system is similar to the Italian one, because it is devolved to the jurisdiction of the administrative judge [15, 28].

4 Approach

Our approach consists of three main steps, as represented in Figure 1. First, there is a data collection step where legal events of interest for the analysis are collected. Then, the resulting dataset is converted into an event log fulfilling

¹¹ <https://www.gazzettaufficiale.it>

¹² <https://ted.europa.eu/TED/main/HomePage.do>

¹³ <https://ec.europa.eu/eurostat/web/nuts/background>

¹⁴ Articles 48, 49, 50, 51 of the directive 2014/24/EU

¹⁵ Directive 2007/66/EC

the basic requirements for the application of PM techniques. It is possible to explore the log to validate the process, identify the main features, exclude any anomalies in the data, and focus on the events and cases of greatest interest. Finally, there is the PM phase, which includes process discovery and variant analysis. As already mentioned in Section 3), our study can be reflected on the various EU calls for tenders as the features collected through ANAC represent a specialisation of the TED ones.

Each step is further detailed in the following subsections.

4.1 Preprocessing and event log creation

The typical main basic step in a PM search is the construction of the log file that includes the time sequence of events. Data obtained from ANAC and IAJ were discussed with domain experts (interviewed several times during the course of the research). Data considered relevant according to domain experts have been used to build the event log related to our legal process. Each event in an event log includes at least three basic features: the identifier of the process it belongs to, the name of the event which generated the event, and the corresponding execution timestamp [2]. In our case, CIG is the identifier, while the different activities involved in the process are 12 (see above), starting from procurement creation, the commencement of work, the suspension of a contract, the arrival of a complaint, up to the end of the contract. Each event includes a date at the level of granularity of the day on which the event occurred. In addition, we add in the log file as features the information on the corresponding region, the sector, the amount, the presence of a sentence.

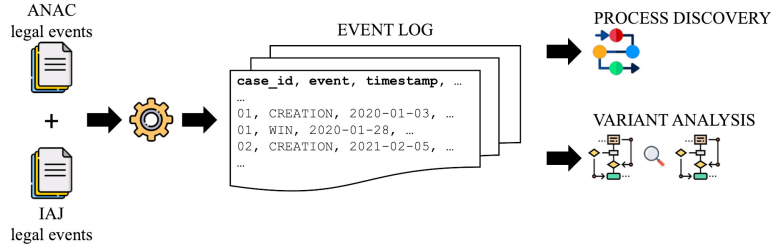


Fig. 1. Methodological steps in our approach from dataset features to event log. Full size image available at <https://github.com/roberto-nai/KDPM4LAW2023>

In order to have a consistent event log, we removed cases too short, or with few activities, which are not meaningful according to domain experts. We then selected cases with at least 5 events: creation, publication, win, contract start, contract end. Table 1 describes the mapping between dataset features and event log concept.

Table 1. Procurement data mapped to event log

Feature	Description	Event log
CIG	Procurement identifier	Case ID
Status	Status of the procurement	Event activity
Status date	Status date of the procurement	Event timestamp
Region	Regional section in which the procurement is published	Resource
Sector	Values: “supplies”, “services”, or “works”	Case feature
Amount	Amount of the procurement	Case feature
Complaint	Presence of a litigation related to the case	Case feature

4.2 Process mining techniques

Based on the event log, we propose different PM analyses. To answer RQ1, discovery algorithms can be applied to automatically derive process models. In the wide range of discovery methods proposed in the literature, we focus on the Fuzzy Miner implementation [17]. This is one of the most used process discovery techniques for real-life, complex processes, thanks to its capability of abstracting from noisy behaviors and returning only the most frequent ones. As a proof of concept for PM, the results were analyzed considering each of the following PM *perspectives*: control-flow, organizational (resource), and time [1]. The control-flow perspective is a type of analysis that focuses on the discovery of the activities’ sequence in the (legal) process. The organizational perspective typically focuses on discovering the resource allocation, the main roles in the organization, and its structure in terms of social network. Finally, the time perspective investigates the occurrence time and frequencies of events to identify bottlenecks in the flow.

Furthermore, we perform variant analysis to discover possible significant differences between subgroups of process executions. Variant analysis is a family of techniques which analyze event logs to identify and explain the differences between two or more kinds of processes. To answer RQ2, in this work we carried out a *manual* comparison of the process models obtained for the identified variants. This is an exploratory step that allows us to gather useful insights into whether these variants actually show some interesting differences and to determine which properties we want to explore further. In terms of technology, we imported the initial CSV log files for further analysis in the free and open source tool ProM¹⁶. Furthermore, the corresponding XES¹⁷ event log file has been used for our experiments with the tool DISCO from Fluxicon¹⁸ to perform an initial analysis of the data involving filtering, exploration with process discovery and bottleneck search.

¹⁶ <https://promtools.org/>

¹⁷ XES is the standard format in a tag-based language [3]: <https://xes-standard.org/>

¹⁸ <https://fluxicon.com/disco/>

5 Results

5.1 Legal event log

The final event log, processed from the dataset as described in Section 4.1, contains data on 43,407 procurement cases and 344,845 occurrences of events with a global mean case duration of 14.8 months. In accordance with Section 3, the event log contains a block of events with a higher occurrence (STATUS AWARD, CREATION, SUSPENSION, CONTRACT-END, CONTRACT-START, from 10% to 19%) and a second block with a lower occurrence (REPRISE, VARIANT, SUBCONTRACT, PUBLICATION-IT PUBLICATION-EU, TESTING, RECOURSE, SENTENCE from 9% to 0.5%). The log was then divided according to case *features* (**sector**, **amount** and **complaint**) for variant analysis (Section 5.3). The exploration of the event log was carried out with the analytical tools of the aforementioned PM tools, making it possible to check its consistency with domain experts. Finally, the main legal event log file (in CSV and XES format) is publicly available on our repository¹⁹.

5.2 Process discovery

Control-flow perspective The discovered process model from the event log provides a complete overview of the actual legal process flow (Figure 2). In the process map, the event with the highest frequency, PUBLICATION, is indicated as the starting point. A group of events has higher frequency, where darker rectangles in the diagrams correspond to “standard” events existing in all procurement. A second group of particular and only eventual events occurs with a lower frequency (lighter color in the map), i.e. SUBCONTRACT, VARIANT or SENTENCE. The diagram reflects what domain experts expected, and makes it easy to specify the corresponding frequencies for the branches of contract suspensions and subcontracts. Nevertheless, the diagram is still quite general, and it can be better articulated by breaking down the log by meaningful features, such as distinguishing cases by region, sector, cases whose amount is under/above threshold, as well as cases with complaint or without, as detailed in Section 5.3.

Organizational perspective The organizational perspective analysis allows process maps and metrics to be evaluated considering the resources (people or devices) that execute the process events. In our case, since procurements are organized at the regional level, the “resource” taken into consideration is the region of the municipality that issued the call. The results indicate the importance first of all of Lombardy’s municipality (87,147 cases, i.e. 25.3%), as well as Tuscany’s municipality (31,348 cases, i.e. 9.1%) and Veneto’s municipality (30,38 cases, i.e. 8.8%); the mean case duration for the municipalities of this three regions is about 14 months.

¹⁹ <https://github.com/roberto-nai/KDPM4LAW2023>

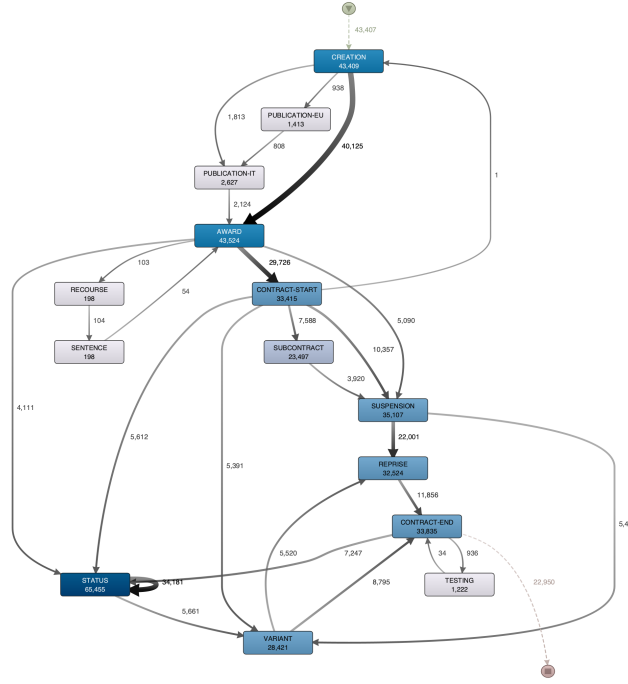


Fig. 2. Process map based on absolute frequency metrics for the entire dataset. Full size image available at <https://github.com/roberto-nai/KDPM4LAW2023>

Time perspective The diagram showing the average duration of transactions between events makes it possible to identify bottlenecks. As highlighted by thicker arcs in Figure 3, consistently with domain experts, the main critical transitions are represented by:

- procurement CREATION to AWARD (about 19 weeks);
- procurement SENTENCE to AWARD (22 weeks on average);
- procurement CONTRACT-START to VARIANT (24.7 weeks on average).

These results are generally in line with what was expected, although the visualisation (improved with animation, in the tools used) made it more immediate and easier to domain experts where critical issues exist and where bottlenecks occur.

5.3 Variant analysis

By discussing with domain experts, we identified three main properties that are of interest for variant analysis, i.e., the case features **sector**, **amount** and **complaint**. We analyzed process variants by taking into account both their performance, expressed in terms of their duration, and their process structure.

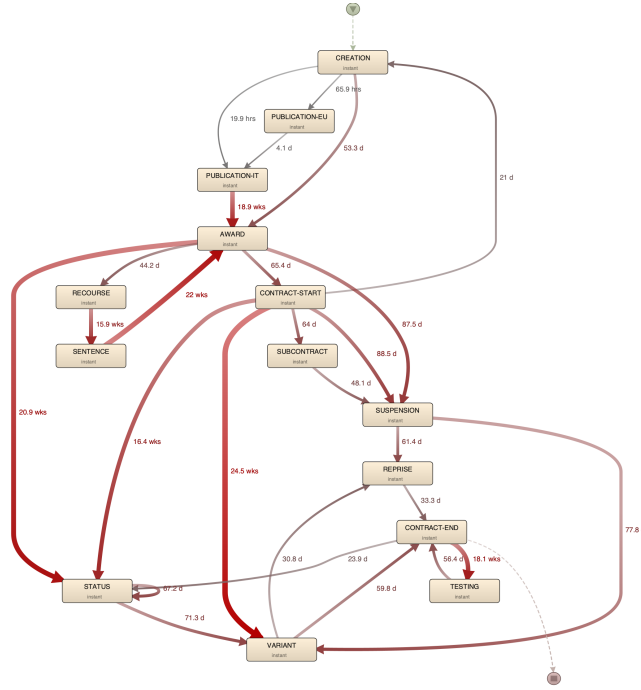


Fig. 3. Time perspective highlighting slow transitions and bottlenecks in some activities of the legal process. Full size image available at <https://github.com/roberto-nai/KDPM4LAW2023>

Below we discuss the main findings we derived by applying a manual variant comparison on a subset of variants considered of special interest.

Manual variant comparison

Process by sector While the global mean case duration is 14.8 months, we focused on the **sector** feature for each case. Therefore, we notice how times by sector of procurement in Table 2 vary by about eleven months (from a minimum of 13 months for “works” to a maximum of 24 months for “services”).

Process above and under a threshold Table 2 describes the differences in terms of case duration when defining variants for the **amount** feature. We used a threshold defined by European law to distinguish between high-amount cases and low-amount ones; we then checked the map of the process for each variant of each sector. For “works” the mean duration of cases with amount under the threshold (30,710 cases) is about 11 months against 21.5 month for cases with amount above the threshold (5,081 cases); for “services” the mean duration of cases with amount under the threshold (2,349 cases) is about 19.4 months against 26.3

month for cases with amount above the threshold (4,878 cases); for “supplies”, the mean duration of cases with amount under the threshold (741 cases) is about 12.5 months against 19.5 month for cases with amount above the threshold (881 cases).

Process with or without complaint The last variants we identified are based on the **complaint** feature. For each case, we notice how times of procurement vary by about seventeen months; the mean duration is 14.7 months for cases without sentences while it is 31 months for cases with a complaint (Table 2). As can be seen in Figure 4, the RECOURSE event occurs in the early stages, after the CREATION, PUBLICATION or AWARD. However, when comparing the control flow of the variants, no relevant differences have been discovered (Figure 4).

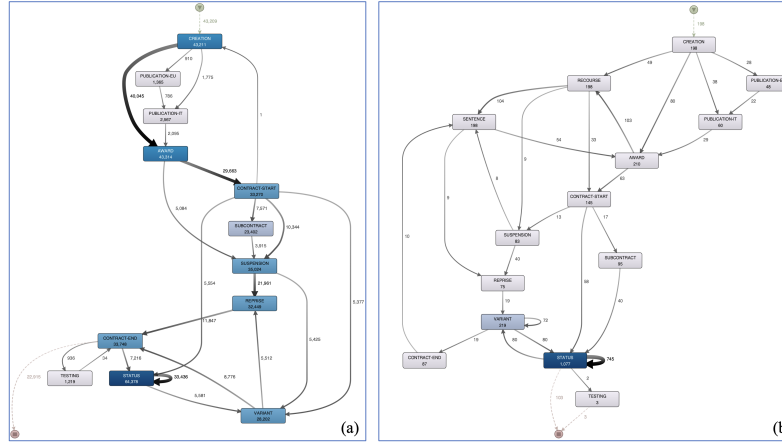


Fig. 4. Process map comparison: (a) procurements without complaint; (b) procurements with a complaint (RECOURSE and SENTENCE events). Full size image available at <https://github.com/roberto-nai/KDPM4LAW2023>

Table 2. Manual variant comparison by case perspective: **sector**, **amount**, **complaint**

Perspective	Value	Mean case duration (months)	Cases
sector	Services	24	48%
	Supplies	16	10%
	Works	13	42%
amount	Under-threshold	Services: 19.4, Supplies: 12.5, Works: 11.3	76%
	Above-threshold	Services: 26.3, Supplies: 19.5, Works: 21.5	24%
complaint	Yes	31	5%
	No	14.7	95%

5.4 Discussion

The current approach has been presented and discussed with domain experts, both from ANAC and legal experts who followed the approach presented in the different stages of the research framework. A first observation concerns the importance for a public body to receive real information from the available data saved in legal information systems. This awareness of having information on calls for tender and the text of tenders was a source of great interest. In some cases, domain experts have shown interest in further developing such analyses by proposing to extend the recording of such data, with the hypothesis of equipping their office in the future with appropriate process-aware information systems, suitable for capturing information on legal events of interest.

A second line of interest concerned the possibility of exploiting the textual data contained in the tenders, by means of NLP techniques. The need to automatically extract information not immediately captured by information systems, such as subcontracting companies, emerged as relevant. Automatic extraction systems such as entity extraction or named entity recognition proved to be of great interest due to the possibility of facilitating the work of legal domain experts.

A third advantage of the proposed method concerns the possibility of comparing different legal processes, distinguishing them on the basis of interest characteristics.

Firstly, variant analysis was useful to distinguish among cases that showed a particularly long duration and those that were relatively short or normal. In addition, it was possible to identify trials that involved particularly large amounts of money. Of similar interest was the possibility of applying process discovery to trials by territorial scope, thus comparing the courts of the most efficient regions. A critical point was made, however, regarding the possibility of comparing trials depending on the judge who conducted the case. Comparing judges is underappreciated because of the possibility of such techniques being perceived as tools to control the work of people and the corresponding staff, effectively limiting their autonomy. On one side, a slower judge is not necessarily adverse, considering the accuracy of the legal analysis and interpretation undertaken. Also, caution is expressed from the point of view of worker protection, for fear of such tools being seen as instruments of control over the work of the individual judge. Finally, a problematic point was found to be the anonymization of data, as there may be references to individuals in the texts of the tenders that cannot be made public. Use for research purposes only is certainly tolerated, but making the dataset available to the scientific community requires an additional anonymization effort that cannot be done from legal staff, who are already overburdened with normal daily activities. Anonymization techniques seem to be a good compromise to facilitate the effort making the datasets usable by the scientific community.

6 Conclusions and future work

In this work, we applied PM techniques to a legal process to understand how this process is executed in reality and to identify potential issues and inefficiencies.

cies. More precisely, we implemented a combination of process discovery and variant analysis techniques. Responding to RQ1, discovery techniques allowed us to gain relevant insights into the main process behaviours. We also identified interesting process variants that deviate significantly from the main process flow. In connection with RQ2, our results demonstrate the importance of taking the heterogeneity of legal processes into account, to be able to implement strategies for process management and improvement tailored to different variants. In the future, we intend to develop our investigation further by exploring different process discovery and variant analysis techniques (e.g: automated variant analysis). Furthermore, we intend to investigate also the feasibility of online PM approaches. For example, we want to study the application of prediction algorithms to detect fallacious or excessively delayed cases, also based on the characteristics of the initial events recorded in the event log.

Acknowledgement

This research has been partially carried out within the “Circular Health for Industry” project, funded by “Compagnia di San Paolo” under the call “Intelligenza Artificiale, uomo e società”.

References

1. van der Aalst, W.M.P.: Process Mining - Data Science in Action, Second Edition. Springer (2016). <https://doi.org/10.1007/978-3-662-49851-4>, <https://doi.org/10.1007/978-3-662-49851-4>
2. van der Aalst, W.M.P., Carmona, J. (eds.): Process Mining Handbook, Lecture Notes in Business Information Processing, vol. 448. Springer (2022). <https://doi.org/10.1007/978-3-031-08848-3>, <https://doi.org/10.1007/978-3-031-08848-3>
3. Acampora, G., Vitiello, A., Di Stefano, B., van der Aalst, W., Gunther, C., Verbeek, E.: Ieee 1849: The xes standard: The second ieee standard sponsored by ieee computational intelligence society [society briefs]. IEEE Computational Intelligence Magazine **12**(2), 4–8 (2017). <https://doi.org/10.1109/MCI.2017.2670420>
4. Amantea, I.A., Robaldo, L., Sulis, E., Boella, G., Governatori, G.: Semi-automated checking for regulatory compliance in e-health. In: 25th International Enterprise Distributed Object Computing Workshop, EDOC Workshop 2021, Gold Coast, Australia, October 25-29, 2021. pp. 318–325. IEEE (2021). <https://doi.org/10.1109/EDOCW52865.2021.00063>, <https://doi.org/10.1109/EDOCW52865.2021.00063>
5. Bogarín, A., Cerezo, R., Romero, C.: A survey on educational process mining. WIREs Data Mining Knowl. Discov. **8**(1) (2018). <https://doi.org/10.1002/widm.1230>, <https://doi.org/10.1002/widm.1230>
6. Bolt, A., de Leoni, M., van der Aalst, W.M.: Process variant comparison: using event logs to detect differences in behavior and business rules. Information Systems **74**, 53–66 (2018)
7. Carlucci, C., Giorgiantonio, C., Orlando, T.: Tempi di realizzazione delle opere pubbliche e loro determinanti, Questioni di economia e finanza. No. 538, Banca d’Italia (2019)

8. Carmona, J., van Dongen, B.F., Solti, A., Weidlich, M.: Conformance Checking - Relating Processes and Models. Springer (2018). <https://doi.org/10.1007/978-3-319-99414-7>, <https://doi.org/10.1007/978-3-319-99414-7>
9. Cavallo Perin, R., Galetta, D.-U., et al.: Il diritto dell'amministrazione pubblica digitale. Giappichelli, Torino (2020)
10. Cavallo Perin, R., et al.: L'amministrazione pubblica con i big data: da Torino un dibattito sull'intelligenza artificiale. Università degli Studi di Torino (2021)
11. Consiglio di Stato - Ufficio Studi: Analisi di impatto del contenzioso amministrativo in materia di appalti (2015-2016), https://www.astrid-online.it/static/upload/cds_/cds_contenzioso-appalti_nsiga_4581622.pdf, last accessed May 2023
12. Consiglio di Stato - Ufficio Studi: Analisi di impatto del contenzioso amministrativo in materia di appalti (2018-2019), <https://www.anticorruzione.it/documents/91439/171948/Consiglio+di+Stato+-+Analisi+di+impatto+del+contenzioso+amministrativo+in+materia+di+appalti+-+biennio+2018+2019.pdf/347b4041-bdc5-d9f6-070d-12a2c8b6a60c?t=1583925605703>, last accessed May 2023
13. Cordes, C., Vogelgesang, T., Appelrath, H.J.: A generic approach for calculating and visualizing differences between process models in multidimensional process mining. In: Business Process Management Workshops: BPM 2014 International Workshops, Eindhoven, The Netherlands, September 7-8, 2014, Revised Papers 12. pp. 383–394. Springer (2015)
14. Fahrenkrog-Petersen, S.A.: Providing privacy guarantees in process mining. In: Rosa, M.L., Plebani, P., Reichert, M. (eds.) Proceedings of the Doctoral Consortium Papers Presented at the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019), Rome, Italy, June 3-7, 2019. CEUR Workshop Proceedings, vol. 2370, pp. 23–30. CEUR-WS.org (2019), <http://ceur-ws.org/Vol-2370/paper-03.pdf>
15. Folliot-Lalliot, L., Torricelli, S., et al.: Contrôles et contentieux des contrats publics = oversight and challenges of public contracts. Bruylant, Bruxelles (2017)
16. Gnaldi, M., Ponti, B.: Misurare la corruzione oggi: Obiettivi, metodi, esperienze (2018)
17. Günther, C.W., van der Aalst, W.M.P.: Fuzzy mining - adaptive process simplification based on multi-perspective metrics. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings. Lecture Notes in Computer Science, vol. 4714, pp. 328–343. Springer (2007). https://doi.org/10.1007/978-3-540-75183-0_24
18. Hashmi, M., Governatori, G., Lam, H., Wynn, M.T.: Are we done with business process compliance: state of the art and challenges ahead. Knowl. Inf. Syst. **57**(1), 79–133 (2018). <https://doi.org/10.1007/s10115-017-1142-1>, <https://doi.org/10.1007/s10115-017-1142-1>
19. Hod, S., Chagal-Feferkorn, K., Elkin-Koren, N., Gal, A.: Data science meets law. learning responsible ai together. Communications of the ACM **2**, 35–39 (2022)
20. Jiang, J., Aldewereld, H., Dignum, V., Wang, S., Baida, Z.: Regulatory compliance of business processes. AI Soc. **30**(3), 393–402 (2015). <https://doi.org/10.1007/s00146-014-0536-9>, <https://doi.org/10.1007/s00146-014-0536-9>
21. Kriglstein, S., Wallner, G., Rinderle-Ma, S.: A visualization approach for difference analysis of process models and instance traffic. In: Business Process Management:

- 11th International Conference, BPM 2013, Beijing, China, August 26-30, 2013. Proceedings. pp. 219–226. Springer (2013)
22. Mannhardt, F., Petersen, S.A., Oliveira, M.F.: Privacy challenges for process mining in human-centered industrial environments. In: 2018 14th International Conference on Intelligent Environments (IE). pp. 64–71 (2018). <https://doi.org/10.1109/IE.2018.00017>
23. Meo, R., Nai, R., Sulis, E.: Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable AI... what's next? In: Silvia Chiusano et al. (ed.) ADBIS 2022, Turin, Italy, September 5-8, 2022. LNCS, vol. 13389, pp. 25–34. Springer (2022). https://doi.org/10.1007/978-3-031-15740-0_3
24. Munoz-Gama, J., Martin, N., Fernández-Llatas, C., Johnson, O.A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., Amantea, I.A., Andrews, R., Arias, M., Beerepoot, I., Benevento, E., Burattin, A., Capurro, D., Carmona, J., Comuzzi, M., Dalmas, B., de la Fuente, R., Francescomarino, C.D., Ciccio, C.D., Gatta, R., Ghidini, C., Gonzalez-Lopez, F., Ibáñez-Sánchez, G., Klasky, H.B., Kurniati, A.P., Lu, X., Mannhardt, F., Mans, R., Marcos, M., de Carvalho, R.M., Pegoraro, M., Poon, S.K., Pufahl, L., Reijers, H.A., Remy, S., Rinderle-Ma, S., Sacchi, L., Seoane, F., Song, M., Stefanini, A., Sulis, E., ter Hofstede, A.H.M., Toussaint, P.J., Traver, V., Valero-Ramon, Z., van de Weerd, I., van der Aalst, W.M.P., Vanwersch, R.J.B., Weske, M., Wynn, M.T., Zerbato, F.: Process mining for healthcare: Characteristics and challenges. *J. Biomed. Informatics* **127**, 103994 (2022). <https://doi.org/10.1016/j.jbi.2022.103994>, <https://doi.org/10.1016/j.jbi.2022.103994>
25. Nai, R., Sulis, E., Marengo, E., Vinai, M., Capecchi, S.: Process mining on students' web learning traces: A case study with an ethnographic analysis. In: Olga Viberg et al. (ed.) EC-TEL Int.Conf.proc. Lecture Notes in Computer Science, vol. 14200, pp. 599–604. Springer (2023). https://doi.org/10.1007/978-3-031-42682-7_48
26. Nai, R., Sulis, E., Meo, R.: Public procurement fraud detection and artificial intelligence techniques: a literature review. In: Symeonidou, D., Yu, R., Ceolin, D., Poveda-Villalón, M., Audrito, D., Caro, L.D., Grasso, F., Nai, R., Sulis, E., Ekaputra, F.J., Kutz, O., Troquard, N. (eds.) Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management, Bozen-Bolzano, Italy, September 26-29, 2022. CEUR Workshop Proceedings, vol. 3256. CEUR-WS.org (2022), <http://ceur-ws.org/Vol-3256/km41aw4.pdf>
27. Nai, R., Sulis, E., Pasteris, P., Giunta, M., Meo, R.: Exploitation and merge of information sources for public procurement improvement. In: Irena Koprinska et al. (ed.) Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Commun. Comput. Inf. Sci., vol. 1752, pp. 89–102. Springer (2022). https://doi.org/10.1007/978-3-031-23618-1_6
28. OECD: Public procurement review procedures (30) (2000). <https://doi.org/https://doi.org/https://doi.org/10.1787/5kml60w0qbvf-en>, <https://www.oecd-ilibrary.org/content/paper/5kml60w0qbvf-en>
29. Pajno, A., Donati, F., Perrucci, A.: Intelligenza artificiale e diritto: una rivoluzione? volume 2. amministrazione, responsabilità e giurisdizione (2022)
30. Pufahl, L., Rehse, J.: Conformance checking with regulations - A research agenda. In: Koschmider, A., Michael, J. (eds.) 11th International Workshop on Enterprise Modeling and Information Systems Architectures, Kiel, Germany, May 21-22, 2021. CEUR Workshop Proceedings, vol. 2867, pp. 24–29. CEUR-WS.org (2021), <http://ceur-ws.org/Vol-2867/paper4.pdf>

31. Rabuzin, K., Modrušan, N., Križanić, S., Kelemen, R.: Process mining in public procurement in croatia. In: Lalic, B., Gracanin, D., Tasic, N., Simeunović, N. (eds.) *Proceedings on 18th International Conference on Industrial Systems – IS’20*. pp. 473–480. Springer International Publishing, Cham (2022)
32. Racca, G.M.: La digitalizzazione necessaria dei contratti pubblici: per un’amazon pubblica. *DPCE online* **45**(4) (2021)
33. Sadiq, S.W., Governatori, G.: Managing regulatory compliance in business processes. In: vom Brocke, J., Rosemann, M. (eds.) *Handbook on Business Process Management 2, Strategic Alignment, Governance, People and Culture*, 2nd Ed, pp. 265–288. International Handbooks on Information Systems, Springer (2015). https://doi.org/10.1007/978-3-642-45103-4_11, https://doi.org/10.1007/978-3-642-45103-4_11
34. Sangil, M.J.: Heuristics-based process mining on extracted philippine public procurement event logs. In: *2020 7th International Conference on Behavioural and Social Computing (BESC)*. pp. 1–4 (2020). <https://doi.org/10.1109/BESC51023.2020.9348306>
35. Sbicca, F., Marino, G.: *Efficienza dei contratti pubblici e sviluppo di indicatori di rischio corruttivo*. ANAC, Roma (2018)
36. Swinnen, J., Depaire, B., Jans, M.J., Vanhoof, K.: A process deviation analysis—a case study. In: *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9*. pp. 87–98. Springer (2012)
37. Taymouri, F., Rosa, M.L., Dumas, M., Maggi, F.M.: Business process variant analysis: Survey and classification. *Knowledge-Based Systems* **211**, 106557 (2021). <https://doi.org/https://doi.org/10.1016/j.knosys.2020.106557>, <https://www.sciencedirect.com/science/article/pii/S0950705120306869>
38. Unger, A.J., Neto, J.F.d.S., Fantinato, M., Peres, S.M., Trecenti, J., Hirota, R.: Process mining-enabled jurimetrics: Analysis of a brazilian court’s judicial performance in the business law processing. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*. p. 240–244. ICAIL ’21, Association for Computing Machinery, New York, NY, USA (2021). <https://doi.org/10.1145/3462757.3466137>, <https://doi.org/10.1145/3462757.3466137>
39. Van Beest, N.R., Dumas, M., García-Bañuelos, L., La Rosa, M.: Log delta analysis: Interpretable differencing of business process event logs. In: *Business Process Management: 13th International Conference, BPM 2015, Innsbruck, Austria, August 31–September 3, 2015, Proceedings 13*. pp. 386–405. Springer (2015)
40. Yin, R.K.: *Case study research: Design and methods*, vol. 5. sage (2009)
41. Zerbino, P., Stefanini, A., Aloini, D.: Process science in action: A literature review on process mining in business management. *Technological Forecasting and Social Change* **172**, 121021 (2021). <https://doi.org/https://doi.org/10.1016/j.techfore.2021.121021>, <https://www.sciencedirect.com/science/article/pii/S0040162521004534>