Towards visualizing and analysing legal proceedings with process mining

Vittoria **Caponecchia**[1], Bernardo **D'Agostino**[1], Giovanni **Comandè**[1], Daniele **Licari**[1] and Andrea **Vandin**[1,2,*]

[1]*Sant'Anna School of Advanced Studies Pisa, Italy*
[2]*DTU Technical University of Denmark, Lyngby, Denmark*

Abstract

Process mining (PM) is a family of data-driven techniques which use data with the goal of studying the process behind the data, i.e., the data-generating process. Despite initially tailored for the engineering and industrial domain, it is becoming popular also in more human-centric domains like the legal and healthcare ones. This paper proposes preliminary steps towards a general-purpose process mining methodology utilizing Fluxicon's Disco tool aimed at analyzing and optimizing the complex processes underlying legal decision making by Courts. We consider specifically the domain of civil proceedings, with a focus on divorce cases. In PM terms, a case is a legal proceeding, and activities are the different internal phases in which a legal case transits from its beginning to the final judgment. The studied process is, therefore, the internal process followed by the Court, possibly varying over the years, to handle specific types of proceedings. By leveraging process mining techniques, this preliminary study examines the evolution of divorce proceedings within a selected Italian court in the time frame 2013-2019, identifying key performance indicators and uncovering hidden process inefficiencies and efficiencies. The findings highlight the potential of process mining to reveal critical process patterns, enabling organizations to make data-driven decisions and implement targeted process improvements.

1. Introduction

Process Mining (PM) is a family of data-driven techniques studying the processes that generated the considered data, i.e., the data generation process [1]. PM enables for the monitoring, enhancement and validation of the actual process underlying a system. Despite initially developed for engineering and industrial domains, PM is now gaining traction in human-centric fields such as the legal and healthcare sectors (see, e.g., [2, 3, 4, 5, 6]). This paper explores the potential of applying Process Mining to Italian legal proceedings. In particular, we focus on joint divorce cases from a selected Italian Court between 2013 and 2021. The dataset has been built by downloading data of proceedings from the "Portale dei Servizi Telematici del Ministero della Giustizia" (portal of telematic services of the italian Ministry of Justice)[1]. By leveraging the Disco Fluxicon tool [7], we aim at visualizing and analyzing these proceedings, and in particular how they are handled (the process that the Court follows to handle them). All mining tasks were done using on the Fuzzy Miner algorithm as available in Disco version 3.3.7.

We remark that we see legal proceedings as instances of processes. Therefore, in some sense, we see Courts as business organizations. We believe that the use of PM in this field could help addressing challenges related to resource management within Courts, often arising from a shortage of personnel. Without compromising the quality of justice or the right to a fair trial merely to reduce processing times,

---

✉ v.caponecchia@santannapisa.it (V. Caponecchia); b.dagostino@santannapisa.it (B. D'Agostino); g.comande@santannapisa.it (G. Comandè); daniele.licari@santannapisa.it (D. Licari); a.vandin@santannapisa.it (A. Vandin)

🆔 0009-0004-2411-7489 (V. Caponecchia); 0009-0004-2411-7489 (B. D'Agostino); 0000-0003-2012-7415 (G. Comandè); 0000-0002-2963-9233 (D. Licari); 0000-0002-2606-7241 (A. Vandin)

[1]We obtained the original data from https://pst.giustizia.it/PST/it/services.page. Due to copyright issues, it is not possible to provide copies of the dataset or of refinements of it.

process mining can streamline procedures by identifying inefficiencies in legal workflows, enabling a more effective analysis and resolution of issues.

The collected data made it possible to reconstruct the activities and processes of these proceedings. By comparing proceedings from different years, we identified, from a procedural point of view, how bottlenecks present in 2013 were addressed in subsequent years. This preliminary work demonstrates the potential of PM in enhancing the efficiency of legal processes, and therefore of the judicial system in general, paving the way for future research and proposals in this area.

To the best of our knowledge, this is one of the first approaches analysing such a dataset using process mining. This paper presents preliminary results in this direction, paving the way for a rich stream of proposals aiming at improving the efficiency of the legal sector. Many works, focused on extracting processes from normative documents, like laws, with complementary goals and challenges (see, e.g., [3, 8]). Another example is [9], which applies PM, in particular the EverFlow [2] tool, to a dataset of business lawsuits from the Court of Justice of the State of Sao Paulo, Brazil. The paper closest to ours in terms of considered data is [10]. Similarly to us, the authors use event-based data on the Italian judicial system. However, the authors consider complementary analyses (as well as different PM tools, namely Apromore[3]). Indeed, PM techniques are used only for an analysis of variants like ours, while the remaining, very interesting, analyses are done using machine learning tools. For example, differently from us, they do not use PM neither to compare proceedings from different years nor to study bottlenecks or reasons for improved performance. Furthermore, they focus on a different type of court, the Court of Appeal of Milan. The proceedings considered refer to civil litigation in general filed before the Court of Appeal (second instance), rather than being limited to joint divorces filed before the Court (first instance).

2. Domain and case study

In this section, we overview the expected structure of the *process underlying* the proceedings from the reference domain. We will discuss at a high level the phase of divorces, namely joint divorces in Italy, which is the case study of this paper.

2.1. Divorces

Divorces may be as well either joint or judicial: in the first case, there is an agreement between the spouses on the conditions to be adopted, and the proceeding will be filed jointly; in the second case, when there is no agreement, the proceeeding may be filed by either spouse. In this work, we will focus on joint divorces, which we will simply call 'divorces'[4]. Fig. 1 graphically depicts the main activities (the internal *statuses* in which proceedings evolve) in a joint divorce. The labels of the boxes contain the names of the activities in Italian. We will discuss the other numerical information in the figure in the next sections. The request for a divorce is filed

Figure 1: A joint divorce proceeding. Labels, color and width of edges denote the time among source and target activity.

---

in to the Court by means of a writ of summon (ɪꜱᴄʀɪᴢɪᴏɴᴇ ʀᴜᴏʟᴏ ɢᴇɴᴇʀᴀʟᴇ, i.e. ʀᴇɢɪꜱᴛʀᴀᴛɪᴏɴ ɪɴ ᴛʜᴇ ɢᴇɴᴇʀᴀʟ ʀᴇɢɪꜱᴛᴇʀ). The Chancellery will assign the case to a Section of the Court (ᴀꜱꜱᴇɢɴᴀᴢɪᴏɴᴇ ᴀ ꜱᴇᴢɪᴏɴᴇ, i.e. ᴀꜱꜱɪɢɴᴍᴇɴᴛ ᴛᴏ ꜱᴇᴄᴛɪᴏɴ), and, in particular, to a judge of that section (ᴅᴇꜱɪɢɴᴀᴢɪᴏɴᴇ ɢɪᴜᴅɪᴄᴇ, i.e. ᴊᴜᴅɢᴇ ᴅᴇꜱɪɢɴᴀᴛɪᴏɴ) and a date will be set for the first hearing of the spouses before the Court in the designated panel (ꜰɪꜱꜱᴀᴢɪᴏɴᴇ ᴜᴅɪᴇɴᴢᴀ ᴄᴀᴍᴇʀᴀʟᴇ, i.e. ꜱᴄʜᴇᴅᴜʟɪɴɢ ᴏꜰ ᴛʜᴇ ᴘᴀɴᴇʟ ʜᴇᴀʀɪɴɢ).

On the scheduled date, the Court, after hearing the parties and receiving the opinion of the public prosecutor (ʀɪᴛᴏʀɴᴏ ᴀᴛᴛɪ ᴅᴀʟ ᴘᴍ/ᴘɢ, i.e. ʀᴇᴛᴜʀɴ ᴏꜰ ᴅᴏᴄᴜᴍᴇɴᴛꜱ ꜰʀᴏᴍ ᴛʜᴇ ᴘʀᴏꜱᴇᴄᴜᴛᴏʀ), pronounces the divorce by judgment (ɪɴ ᴅᴇᴄɪꜱɪᴏɴᴇ, i.e ᴜɴᴅᴇʀ ᴅᴇᴄɪꜱɪᴏɴ). The judgment is first prepared with its minutes by the reporting judge (ᴅᴇᴘᴏꜱɪᴛᴏ ᴍɪɴᴜᴛᴀ ꜱᴇɴᴛᴇɴᴢᴀ ᴅᴇꜰɪɴɪᴛɪᴠᴀ, i.e. ꜰɪʟɪɴɢ ᴏꜰ ᴛʜᴇ ꜰɪɴᴀʟ ᴊᴜᴅɢᴍᴇɴᴛ ᴅʀᴀꜰᴛ) and submitted to the President of the panel for her/his signature. After this, the Chancellery will take care of the publication (ᴅᴇᴘᴏꜱɪᴛᴏ ꜱᴇɴᴛᴇɴᴢᴀ − ᴘᴜʙʙʟɪᴄᴀᴢɪᴏɴᴇ, i.e. ꜰɪʟɪɴɢ ᴏꜰ ᴊᴜᴅɢᴍᴇɴᴛ ᴀɴᴅ ᴘᴜʙʟɪᴄᴀᴛɪᴏɴ).

At any time during the proceeding, it might be necessary to replace a judge (ꜱᴏꜱᴛɪᴛᴜᴢɪᴏɴᴇ ɢɪᴜᴅɪᴄᴇ, i.e. ᴊᴜᴅɢᴇ ʀᴇᴘʟᴀᴄᴇᴍᴇɴᴛ), for reasons related, for example, to her/his unavailability, resignation, change of duties, etc. This is not shown in Fig. 1 because it is a nonregular event since it does not necessarily arise in all proceedings. Actually, proceedings might have different *variations* of such events, like ꜱᴏꜱᴛɪᴛᴜᴢɪᴏɴᴇ ɢɪᴜᴅɪᴄᴇ ᴘᴇʀ ʟ'ᴜᴅɪᴇɴᴢᴀ (i.e. ʀᴇᴘʟᴀᴄᴇᴍᴇɴᴛ ᴏꜰ ᴊᴜᴅɢᴇ ꜰᴏʀ ᴛʜᴇ ʜᴇᴀʀɪɴɢ), and ꜱᴏꜱᴛɪᴛᴜᴢɪᴏɴᴇ ɢɪᴜᴅɪᴄᴇ ᴇ ꜱᴇᴢɪᴏɴᴇ (i.e. ʀᴇᴘʟᴀᴄᴇᴍᴇɴᴛ ᴏꜰ ᴊᴜᴅɢᴇ ᴀɴᴅ ꜱᴇᴄᴛɪᴏɴ). In Section 3.2 we will mention how, in cases like this, we opted for replacing all such similar but different activities with a single simpler one (in this case, just ꜱᴏꜱᴛɪᴛᴜᴢɪᴏɴᴇ ɢɪᴜᴅɪᴄᴇ, i.e. ᴊᴜᴅɢᴇ ʀᴇᴘʟᴀᴄᴇᴍᴇɴᴛ). We did this for the sake of simplicity; yet a more granular analysis would potentially highlight which variation eventually creates which consequence (e.g., bottlenecks). Notably, as we will see in the next sections, this activity might actually be frequent, and a major reason of delays in the proceeding.

3. Dataset description and pre-processing

In this section, we describe our dataset from a computational and procedural point of view (Section 3.1), we detail the domain-specific pre-processing steps we applied (Section 3.2), and we conduct a preliminary exploratory data analysis comparing two reference years in which the proceedings took place (Section 3.3).

3.1. Dataset description

The dataset has been built by downloading metadata of proceedings from the "Portale dei Servizi Telematici del Ministero della Giustizia" (portal of telematic services of the italian Ministry of Justice)[5]. We considered only a specific Court in northern Italy, focusing on joint divorces within the time-frame 2013-2021. To better identify structural differences in the proceedings across different years, we considered only proceedings starting within the first five months of each year. The obtained dataset contains 68 605 activities like those discussed in Section 2.

In this work, we used the process mining tool Disco from Fluxicon[6] to visualize and analyze the considered legal proceedings. Indeed, Fig. 1 is actually a *map* computed by Disco for a specific proceeding. It shows how one joint divorce proceeding can be visualized in Disco using the so-called *performance view*. The performance view of Disco depicts the activities appearing in the data (in this case the internal statuses of a proceeding), as well as the waiting times between them (the labels of the edges). Boxes represent the activities, and have label *instant* because we only know the day in which a proceeding changed in the corresponding status.

In particular, for each activity of each proceeding, our dataset contains the date when it occurred, the name of the activity, the identifier of the specific proceeding to which the activity belongs, whether it regards a divorce, and whether the legal proceeding is concluded or not at the time in which we

[5] https://pst.giustizia.it/PST/it/services.page
[6] https://fluxicon.com/disco/

downloaded the data. The dataset encompasses 2 560 distinct divorce proceedings and contains 72 different activities.

Feeding to Disco more proceedings, enables process mining tasks, and, in particular, a number of visualization and analysis features. For example, Disco computes a number of statistics on the proceedings, that can be used as alternative label edges (e.g., min/max/average/median times, or counts of proceedings making that state transition). Another example is that Disco can group proceedings in different *variants*, where a variant is a unique sequence of intermediate states of a proceeding. A deeper analysis based on the variants in the dataset is provided in Section 4.2. The proceeding in Fig. 1 belongs to the most common variant for divorces in our dataset (after applying the data cleaning steps discussed next). Considering the temporal information in the map, we can see, e.g., that it changes within the same day from ISCRIZIONE RUOLO GENERALE (i.e. REGISTRATION IN THE GENERAL REGISTER) to ASSEGNAZIONE A SEZIONE (i.e. ASSIGNMENT TO SECTION).[7] Instead, it took 3 days to further change into DESIGNAZIONE GIUDICE (i.e. JUDGE DESIGNATION).

3.2. Domain-driven data cleaning

Before analyzing this dataset, it is necessary to apply a number of *domain-specific* pre-processing steps. Essentially, this can be considered a data-cleaning step. Furthermore, we focused only on concluded proceedings, using the corresponding information in the logs mentioned in Section 3.1. We got 2 300 such divorce proceedings.

Another issue in this dataset is that proceedings have *spurious* activities after their actual conclusion irrelevant to this study. From our domain knowledge, we know that a proceeding can be considered completed after the activity DEPOSITO SENTENZA - PUBBLICAZIONE (i.e FILING OF JUDGMENT AND PUBLICATION) for the divorces. Most of the proceedings in our dataset contained additional activities after those final ones, preventing us from making any reasonable analysis on the duration of proceedings. These are activities such as ANNOTAZIONE (i.e. ANNOTATION), which involve small changes or notes added to the records of the proceedings, or PASSAGGIO IN ARCHIVIO (i.e. TRANSFER TO ARCHIVE), which indicates that the document or proceeding is being transferred to the archive. Based on this understanding, we removed all activities occurring after the "final activities" mentioned above.

As anticipated in Section 2.1, another issue regards groups of activities with different name, but similar meaning. We identified such groups of activities, and consolidated them under a single one. In particular, activities DESIGNAZIONE GIUDICE (i.e. JUDGE DESIGNATION) and DESIGNAZIONE GIUDICE E FISSAZIONE PRIMA UDIENZA (i.e. JUDGE DESIGNATION AND SCHEDULING OF INITIAL HEARING) were merged as DESIGNAZIONE GIUDICE (i.e. JUDGE DESIGNATION). Similarly, the entries SOSTITUZIONE GIUDICE PER L'UDIENZA (i.e. REPLACEMENT OF JUDGE FOR THE HEARING), SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT), and SOSTITUZIONE GIUDICE E SEZIONE (i.e. REPLACEMENT OF JUDGE AND SECTION) were consolidated in SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT).

After these data cleaning steps, we obtained 50 unique activities for the 2 300 divorce proceedings.

We then refined the dataset by focusing on the *relevant* activities. Specifically, we considered 21 unique activities for divorces. In order to select these activities, we identified the 30 most frequent activities in the dataset, we manually analyzed and selected subsets that we deemed representative for the considered proceedings.

Finally, since, as discussed, our dataset only contains the day on which an activity occurred, and not the time and duration, we encountered proceedings where multiple activities took place on the same day, possibly written in a different order proceeding by proceeding. This led to erroneous loops in the mined processes. To resolve this, we defined a predetermined order for the activities, and used it to sort activities happened on the same day for a proceeding. We did this by adjusting the begin time from the default 00:00 by adding as many minutes as the position of the activity in this predefined order. For example, Fig. 1 shows that SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT) happens 60 seconds after DESIGNAZIONE GIUDICE (i.e. JUDGE DESIGNMENT) in the considered proceeding. Indeed, we imposed an ordering where the former is first, and the latter is second (therefore we added 1 minute to its starting time).

---

[7]The label should be *0 days*. We will explain later why we modified the dataset in this way during pre-processing.

Table 1
Duration and number of divorce proceedings across years.

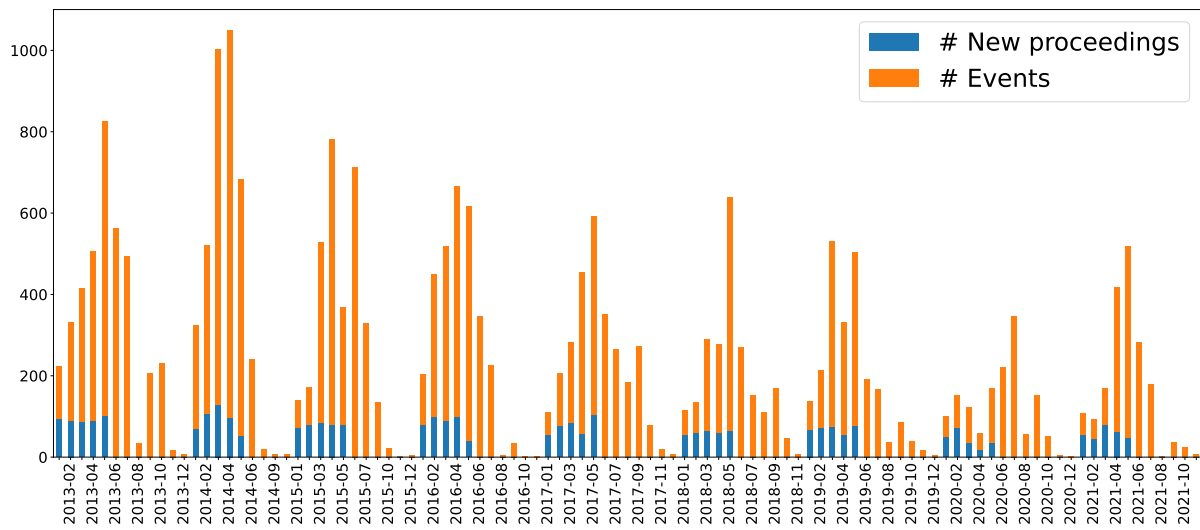| Considered year	Median length in days	Number of divorces
2013	181	263
2014	82	277
2015	159	235
2016	96	357
2017	155	280
2018	120	250
2019	93	238
2020	170	136
2021	140	264

Figure 2: Frequencies of divorce proceedings and of their activities for each month. Due to space constraints, we write only every second x-tick label

After completing all pre-processing and data cleaning steps, we obtained 23 360 events related to 2 300 divorce proceedings.

3.3. Exploratory Data Analysis

Before delving into the process mining capabilities, we present preliminary exploratory data analyses to provide context.

Table 1 summarizes the number of divorce proceedings in the Court for each year along with their median duration. We can see that the median duration fluctuates across the years, and that 2020 was the year with the lowest number of divorces and the second-highest median duration. In particular, in Fig. 2 we can see that the number of new proceedings in the first two months of 2020 was in line with that of previous years, while there has been a significant drop in the following three months. This decline is likely attributable to the COVID-19 outbreak and the corresponding restrictions implemented in Italy starting from the 9th of March. In addition, Table 1 also indicates a return to pre-pandemic trends in terms of the number of proceedings and associated waiting times in 2021, suggesting a mitigation of the pandemic's impact.

4. Towards process mining legal proceedings

This section presents two relevant examples of process mining analyses conducted with Disco to visualize and analyze our data. In Section 4.1, we compare the proceedings belonging to the variant

shown in Fig. 1 with those from a different variant. Then, Section 4.2 offers a preliminary comparison of proceedings started in two years of interest.

4.1. Analysis of two variants

As discussed, Disco can categorize proceedings in variants. For example, we have seen how the proceeding in Fig. 1 belongs to the most common one in the dataset. In Fig. 3a we show an example of a proceeding belonging to the 9th most frequent variant. This latter variant has been chosen for presentation reasons to highlight the capabilities of PM and Disco. This proceeding follows the same path as the one from Fig. 1 up to judge replacement (SOSTITUZIONE GIUDICE). Then, after 12 days, it progresses to the decision phase (IN DECISIONE). One day later, the judge is replaced (SOSTITUZIONE GIUDICE, the upward edge) and she/he is replaced again 20 days later (indicated by the self-loop edge). On the same day, a panel is appointed for the proceeding (DESIGNAZIONE COLLEGIO, i.e. PANEL DESIGNATION) and then the draft of the final judgment is submitted (DEPOSITO MINUTA SENTENZA DEFINITVA). Finally, 3 days later the proceeding terminates with the publication of the judgment (DEPOSITO SENTENZA-PUBBLICAZIONE). In the present analysis, the comparison of the two variants reveals a significant divergence in the path taken by a minority of cases after the JUDGE REPLACEMENT activity. While the majority of cases proceed directly to SCHEDULING OF THE PANEL HEARING, a smaller subset undergoes a series of additional steps, including a return to JUDGE REPLACEMENT and the introduction of a PANEL DESIGNATION activity. This atypical path suggests potential complexities or exceptional circumstances affecting these cases.

In Fig. 3b we show an analysis of all the proceedings in our dataset which follow the two mentioned variants. In particular, we use Disco to mine a map for all such proceedings. Overall, the two variants include 448 proceedings, 418 from the variant considered in Fig. 1, and 30 from the newly considered one.

Differently from Fig. 1, Fig. 3b shows a so-called *frequency view* offered by Disco. Rather than focusing on temporal and performance aspects, it just counts the number of times an activity is met by proceedings (numbers in the boxes), and the number of times proceedings change among two states (labels of edges). The visualization of process frequencies provides a quantitative overview of the most common process flows and their relative frequencies. This information can be used to identify critical control points within the process and to assess the impact of potential process changes. By looking at the top edge, we can see that all proceedings start from ISCRIZIONE RUOLO GENERALE (i.e. REGISTRATION IN THE GENERAL REGISTER). Furthermore, the three following down-stream edges confirm that all 448 proceedings share the same initial path. Then, from SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT), 418 proceedings move to FISSAZIONE UDIENZA CAMERALE (i.e. SCHEDULING OF THE PANEL HEARING). Indeed, all 418 proceedings from the variant in Fig. 1 have such transition, while the 30 from Fig. 3a do not have it. Indeed, the 30 proceedings from the 9th variant jump to activity IN DECISIONE (i.e. UNDER DECISION), get back to SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT) and then move to DESIGNAZIONE COLLEGIO (i.e. PANEL DESIGNATION). The two paths merge back in the activity DEPOSITO MINUTA SENTENZA DEFINITIVA (i.e. FILING OF THE FINAL JUDGMENT DRAFT) and conclude in the same way. By combining process variant analysis with frequency-based insights, domain experts can gain a comprehensive understanding of the underlying process dynamics and identify areas for improvement in the courts.

4.2. PM-based comparison of proceedings from 2013 and 2019

This section employs process mining techniques to perform a preliminary comparison of proceedings from 2013 and 2019, representing the earliest and latest available data points prior to the COVID-19 pandemic.

Quantitative comparison. As shown in Table 1, we have similar numbers of proceedings in the two years, 263 in 2013, and 238 in 2019. However, we find a notable difference: the median duration of proceedings in 2019 is about half that from 2013, denoting sensible improvements in the efficiency of the
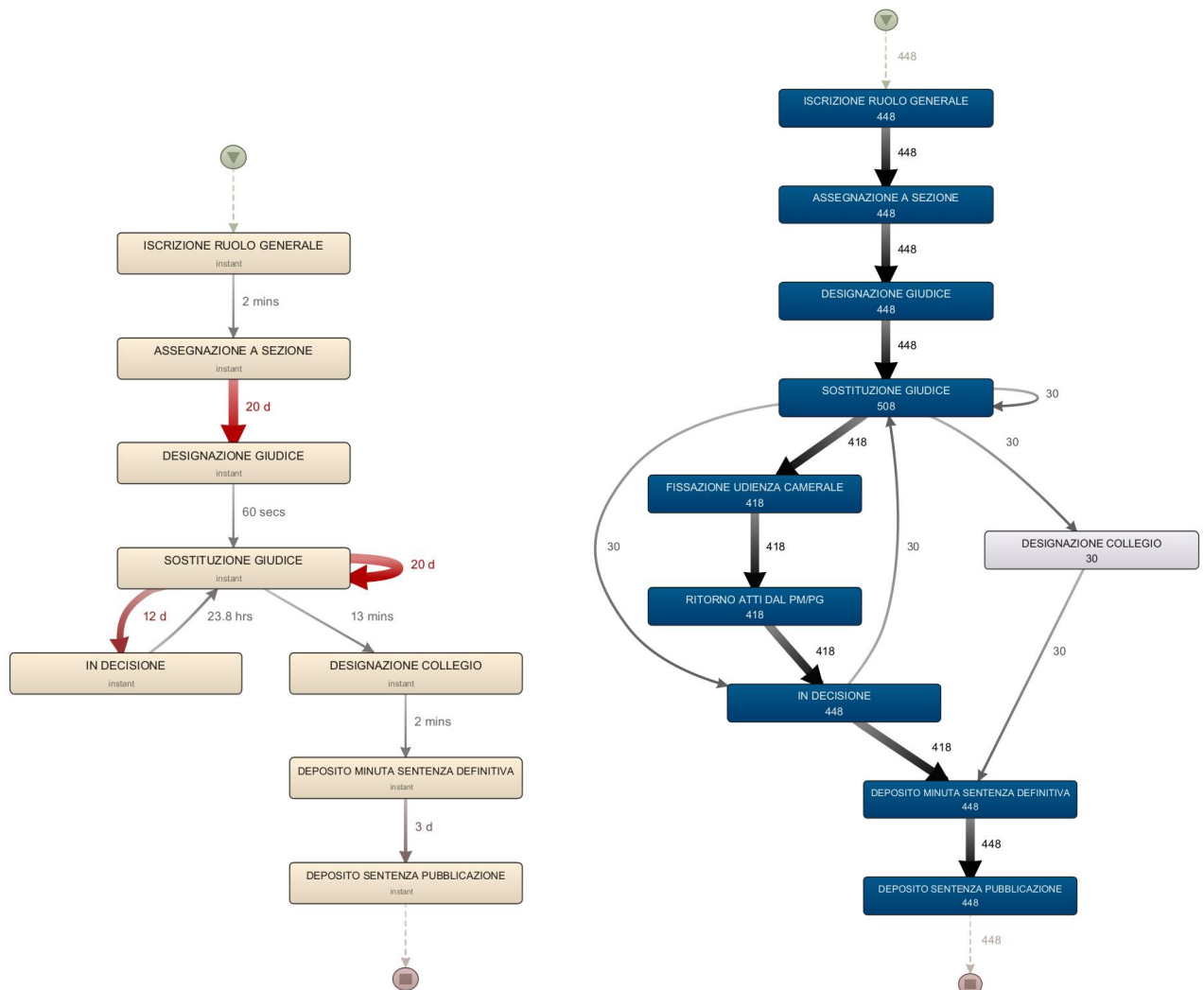
(a) A variant of divorce proceedings

(b) Map mined for variants in Figs. 1 and 3a.

Figure 3: Mining of two variants using Disco. Fig. 3a uses Disco's performance view, while Fig. 3b uses the frequency one.

handling of these proceedings. Then, process mining is employed to shed light on the "organizational and structural reasons" beyond these improvements.

We first analyzed the four most popular variants present in the two years. For doing this, we applied filters in Disco to select only proceedings starting in the first five months of the two years. We found 41 variants in 2019 and 48 in 2013. Despite the number of variants being similar, Table 2 highlights an important aspect: 2019 has a much more homogeneous treatment of proceedings. In fact, the most common variant covers 65% of the proceedings, with the others being below 5%. Instead, there is much less difference among the first three variants from 2013, which together cover a smaller ratio than the top one in 2019 (63%). The comparative analysis of divorce proceedings from 2013 and 2019 reveals a substantial improvement in process efficiency over time. The halving of median processing time in 2019 compared to 2013 underscores a significant transformation in the court's operational practices. Process mining techniques have enabled a deeper understanding of these changes by examining process variations. The emergence of a dominant process variant in 2019, encompassing a significantly larger proportion of cases compared to the previous year, indicates a notable increase in process standardization. This standardization is likely a key factor contributing to the observed efficiency gains.

Table 2
Coverage and Median Time (in days) of Different Variants for 2013 and 2019

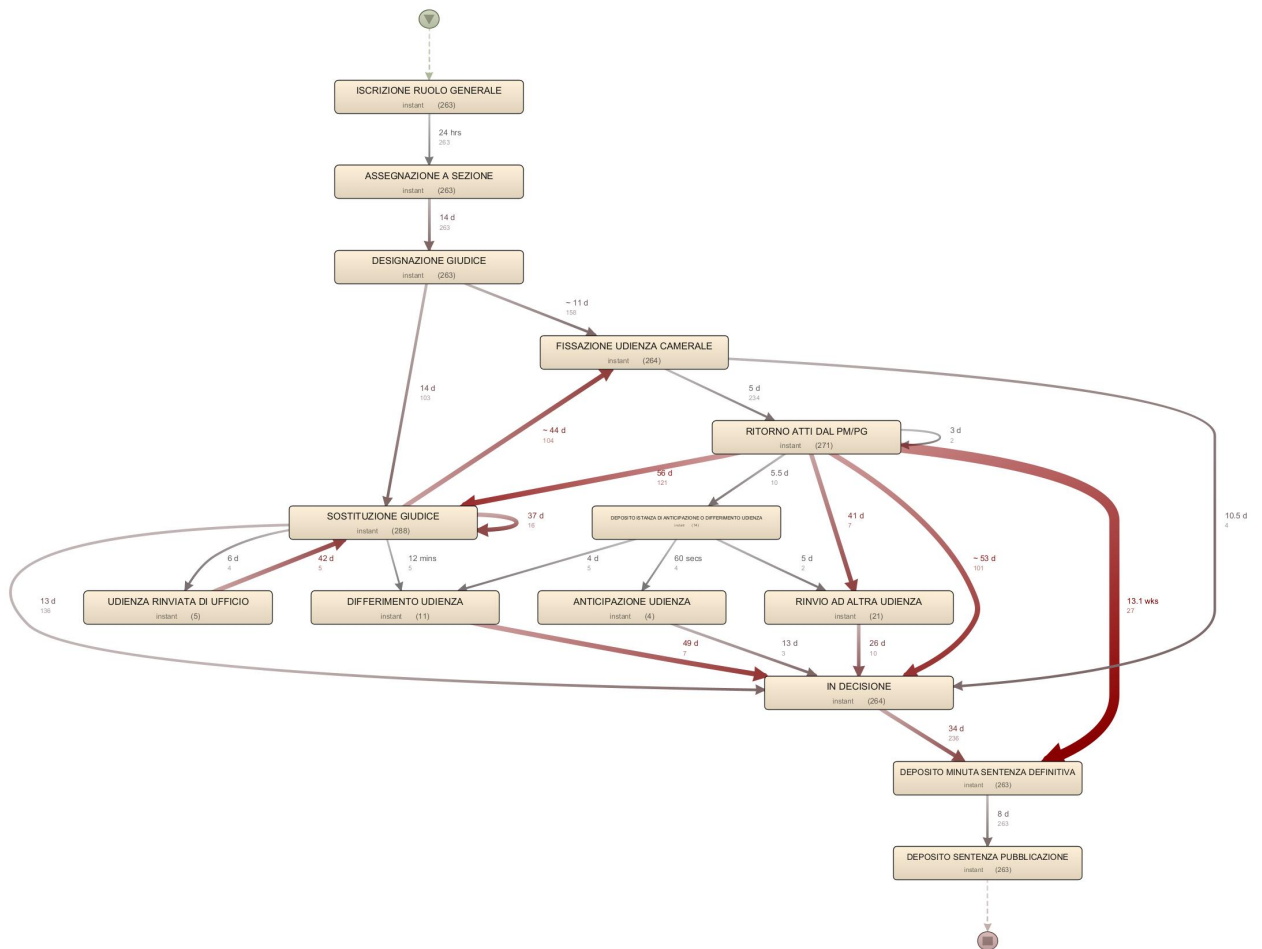| Variant	2013		2019	
	Coverage	*Median Time*	*Coverage*	*Median Time*
1	29%	155	65%	87
2	23%	204	4%	98
3	10%	129	3%	117
4	5%	245	3%	110

Figure 4: Median waiting time and frequencies on the proceedings started in 2013.

Procedural comparison. We then proceeded to analyze graphically the proceedings from the two years. In Figs. 4 and 5 we see the performance views produced by Disco for these two sets of proceedings. The performance view was selected, utilizing the median waiting times as metrics. Notably, Disco offers a *zooming* functionality which allows to focus on more frequent activities and edges.[8] For the maps in Figs. 4 and 5 we decided to use the intermediate setting of 50% for both activities and edges to ignore uncommon edges and activities in this preliminary analysis.

The process maps in Figs. 4 and 5 incorporate secondary labels. Indeed, so far, we focused our performance analyses on the median waiting times among activities. Here we exploit another feature of

---

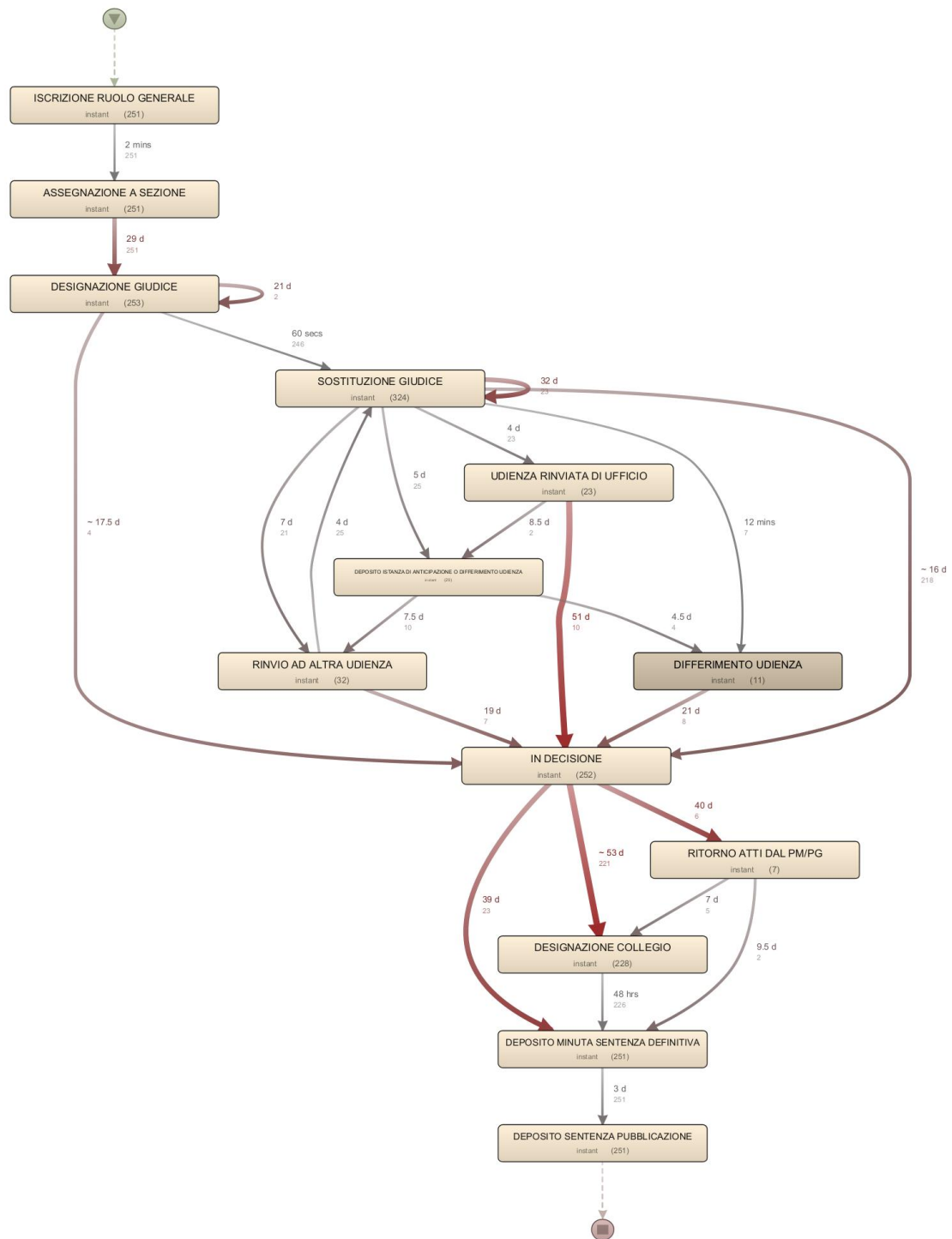[8]See, e.g., https://fluxicon.com/book/read/mapview/

Figure 5: Median waiting time and frequencies on the proceedings started in 2019.

Disco; namely, the ability to add a secondary metrics to a map (the secondary labels). In particular, as secondary metrics we add frequency information, de facto merging the two performance and frequency views. For example, the first activity for the year 2013 (ISCRIZIONE RUOLO GENERALE, i.e. REGISTRATION IN THE GENERAL REGISTER) is found in all 263 proceedings, and the edge outgoing from it is taken by all of them.

Considering aspects in common, we can see that, in both years, all edges outgoing from IN DECISIONE

(i.e. UNDER DECISION) take a considerable amount of time, among the highest ones in the process. This means that the activity IN DECISIONE itself takes a considerable amount of time.

The Italian Code of Civil Procedure (cpc) does not impose formal time limits within which judges must decide a case, except for the labour rite. This is why proceedings often last a long time. Article 275 of the cpc, concerning the decision of the panel, provided [9] that the judgement was to be filed within 60 days of the expiry of the time limit for filing reply briefs or, in the case of oral discussion before the panel, that it was to be signed by the President and immediately filed in the Chancellery. In joint divorce cases, to which the labour rite does not apply, the rules provided for in Article 275 cpc should be applied, since the divorce law does not lay down time limits for the decision. In any case, the 60 days should not be considered to start from the filing of the briefs, activity that in a joint divorce has no reason to take place, but from the discussion. In this way, divorce proceedings would also be subject to time limits and avoid being too lengthy.

Then, a principle that is certainly applicable is that of the reasonable duration of proceedings, dictated by Article 111 of the Constitution [10] and valid for any type of trial. Notably, the secondary metrics tells us that these *time consuming transitions* are particularly important in the overall performance of the Court because they are *frequent*: all proceedings in 2019 involve one of such transitions. Instead, e.g., the transition from UDIENZA RINVIATA DI UFFICIO (i.e. HEARING POSTPONED EX-OFFICIO) to IN DECISIONE (i.e. UNDER DECISION) has a much lower impact because it involves only 10 proceedings in 2019. Also with regard to postponement of hearings, apart from what is established for labour proceedings [11], there are no express prohibitions. However, always following the requirements of Article 111 of the Constitution and the principles of concentration and efficiency governing the trial, there is a tendency not to allow mere postponements except in cases where this is inevitable [12]. This aims to avoid lengthy trials with the consequent waste of money, which is probably why, in 2019, far fewer cases were postponed to a new hearing and moved more quickly to IN DECISION (i.e. IN DECISIONE).

Two key differences emerge from the process maps. First, assigning a judge to a proceeding in 2019 appears to take nearly twice as long compared to 2013. Second, one of the most common activities in the two years, namely SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT), has a much different role in 2019. In 2013, it was central in the process, with incoming edges with very high times. Instead, in 2019 this activity seems to have been moved earlier in the process flow, resulting in a significant reduction in processing time. That is, the incoming edge with a frequency of 246 is instantaneous, while the one with higher waiting times only involves 32 proceedings. These observations suggest an overall improvement in process efficiency, despite the apparent slowdown in the initial judge assignment stage.

Finally, we note how, in 2019, activities DESIGNAZIONE GIUDICE (i.e. JUDGE DESIGNATION) and SOSTITUZIONE GIUDICE (i.e. JUDGE REPLACEMENT) occur on the same day. We remark that real substitutions are very unlikely to happen so frequently within a day. One explanation could be that, e.g., proceedings have an initial temporary automatic assignment, which is then refined and updated. This did not appear in 2013 because this procedure was probably instrumented in a different manner. This analysis demonstrates the effectiveness of process mining in revealing process variations and performance changes over time. By providing a visual and quantitative representation of process execution, Disco facilitates the identification of bottlenecks and areas for improvement, ultimately contributing to more efficient and streamlined legal procedures.

---

[9] The use of the past tense is necessary since the Cartabia Reform amended many provisions of the Code of Civil Procedure, including this one. However, since the reform came into force as of 28 February 2023, the old provision of the article applies to this work.

[10] Already recognised by Article 6 ECHR and Article 47 of the EU Charter of Fundamental Rights.

[11] Article 420 of cpc expressly prohibits mere postponements of hearings, but this does not apply to cognitive proceedings.

[12] E.g., Judgment No. 26935 of October 24, 2018, Supreme Court of Cassation, Civil Section, "Not only in the labour rite... but also in the ordinary rite of cognition mere postponement hearings are not allowed."

5. Conclusions

This preliminary work demonstrates the potential of Process Mining (PM) in monitoring and enhancing the efficiency of the judicial system by mining and reasoning about the process underlying the handling of legal proceedings. We employed process mining techniques to analyze the evolution of divorce proceedings within an Italian Court, focusing on the period between 2013 and 2019. We identified bottlenecks, superfluous activities, and different trends over the years. The findings reveal a significant enhancement in process efficiency over the examined period, characterized by a substantial reduction in median case processing times and an increased degree of process standardization. The emergence of a dominant process variant in 2019 suggests successful implementation of streamlined procedures and optimized resource allocation. Ultimately, this research underscores the potential of process mining, exemplified by tools like Fluxicon Disco, in analyzing and optimizing intricate court operations. By unveiling previously hidden patterns and inefficiencies, these user-friendly tools empower, for instance, courts' administration to make data-driven decisions and spearhead positive transformations within the court system. Future research can build on these findings to support decision makers (e.g., Presidents of Courts) in streamlining legal processes by promoting standardization, efficiency, and data-driven decision-making. Indeed, this is one of the main goals of process mining, and more in general, business process management [1, 11], also in other domains. It also promises to support legal reform by data-driven evidence for identifying sources of relevant problems in the legal process and possible solutions. improve the efficiency of the legal sector. By leveraging the potential of process mining, legal institutions may embark on a data-driven journey towards continuous process improvement and enhanced service delivery. In particular, we leave for future works deeper analyses on the impact of COVID'19 in the proceedings in 2020. Furthermore, deeper domain-specific considerations are necessary: some proceedings simply *have* to take longer due to case-by-case needs, independently of inefficiencies. For example, if the Court finds that the agreement among the spouses is not in the best interests of the offspring, the President of the Court shall nominate an investigating judge to whom the case is referred. Finally, it might be useful to involve the Chanchellery of the considered Court to gain a deeper understanding of certain aspects of the studied process by those who have actually managed these processes.

Acknowledgments

References

[1] W. M. van der Aalst, Process Mining, 2nd ed., Springer, 2016.

[2] Wil Van der Aalst and Roberto Gatta and Carlos Fernandez Llatas and others, Process mining for healthcare, https://www.processmining4healthcare.org/, Access: September 2023.

[3] H. A. López, Challenges in legal process discovery, in: A. Marrella, D. T. Dupré (Eds.), Proceedings of the 1st Italian Forum on Business Process Management co-located with the 19th International Conference of Business Process Management (BPM 2021), Rome, Italy, September 10th, 2021, volume 2952 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 68–73. URL: https://ceur-ws.org/Vol-2952/paper_302a.pdf.

[4] H. A. López, S. Debois, T. Slaats, T. T. Hildebrandt, Business process compliance using reference models of law, in: H. Wehrheim, J. Cabot (Eds.), Fundamental Approaches to Software Engineering - 23rd International Conference, FASE 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25-30, 2020, Proceedings,

volume 12076 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 378–399. URL: https://doi.org/10.1007/978-3-030-45234-6_19. doi:10.1007/978-3-030-45234-6_19.

[5] H. A. López, R. Strømsted, J. Niyodusenga, M. Marquard, Declarative process discovery: Linking process and textual views, in: S. Nurcan, A. Korthaus (Eds.), Intelligent Information Systems - CAiSE Forum 2021, Melbourne, VIC, Australia, June 28 - July 2, 2021, Proceedings, volume 424 of *Lecture Notes in Business Information Processing*, Springer, 2021, pp. 109–117. URL: https://doi.org/10.1007/978-3-030-79108-7_13. doi:10.1007/978-3-030-79108-7_13.

[6] S. Debois, H. A. López, T. Slaats, A. A. Andaloussi, T. T. Hildebrandt, Chain of events: Modular process models for the law, in: B. Dongol, E. Troubitsyna (Eds.), Integrated Formal Methods - 16th International Conference, IFM 2020, Lugano, Switzerland, November 16-20, 2020, Proceedings, volume 12546 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 368–386. URL: https://doi.org/10.1007/978-3-030-63461-2_20. doi:10.1007/978-3-030-63461-2_20.

[7] C. W. Günther, A. Rozinat, Disco: discover your processes, in: proceedings of demonstration track of the 10th international conference on Business Process Management, BPM Demos'12, Tallinn, Estonia, Springer-Verlag, Berlin, Heidelberg, Tallinn, Estonia, 2012, pp. 40–44. https://research.tue.nl/en/publications/disco-discover-your-processes.

[8] D. Licari, G. Comandé, ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the italian legal domain, Comput. Law Secur. Rev. 52 (2024) 105908. URL: https://doi.org/10.1016/j.clsr.2023.105908. doi:10.1016/J.CLSR.2023.105908.

[9] A. J. Unger, J. F. d. S. Neto, M. Fantinato, S. M. Peres, J. Trecenti, R. Hirota, 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, ICAIL '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 240–244. URL: https://doi.org/10.1145/3462757.3466137. doi:10.1145/3462757.3466137.

[10] B. Pernici, M. Dilettis, P. Gerosa, A. Campi, Why are italian trials taking so long? A process mining approach, in: F. Falchi, F. Giannotti, A. Monreale, C. Boldrini, S. Rinzivillo, S. Colantonio (Eds.), Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023, volume 3486 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 284–289. URL: https://ceur-ws.org/Vol-3486/34.pdf.

[11] W. M. P. van der Aalst et al., Process mining manifesto, in: F. Daniel, K. Barkaoui, S. Dustdar (Eds.), Business Process Management Workshops, Revised Selected Papers, volume 99 of *Lecture Notes in Business Information Processing*, Springer, 2011, pp. 169–194. doi:10.1007/978-3-642-28108-2_19.