From Text to Process: Leveraging LLMs to Unveil Italian Lawmaking

Matilde Contestabile[1], Chiara Ferrara[1], Alberto Giovannetti[1], Giovanni Parrillo[1], and Andrea Vandin[1,2]

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

Abstract. The Italian lawmaking process involves a complex interplay of institutions, actors, and procedural stages. Despite the existence of previous research, systematic process-oriented studies remain scarce, largely due to the absence of structured data detailing its procedural steps. We introduce ProLiFIC (Procedural Lawmaking Flow in Italian Chambers), a machine-accessible dataset of Italian legislative preparatory works from 1985 to the present. These texts are available in unstructured form on the Normattiva portal. By leveraging large language models for event extraction, we transform this data into structured event logs suitable for analyses in process-oriented Data Science and Process Mining (PM). Our approach promotes transparency in process studies and paves the way for analyses of procedural dynamics, temporal patterns, and inefficiencies in the Italian lawmaking process. Preliminary exploratory data analyses and PM tasks demonstrate the dataset's potential for supporting inquiries in political and legal studies.

Key words: Normattiva, Lawmaking process, Process mining, LLMs

1 Introduction

The Italian lawmaking process, characterized by a complex interplay of institutions, actors and procedural stages, offers fertile ground for empirical analyses across multiple disciplines, including political science, law, and public administration (see, e.g., [1, 2, 3]). Nevertheless, systematic studies of this process remain constrained by the lack of structured data capturing the detailed progression of lawmaking procedures. While previous research has examined aspects such as party coalitions, party switches, legislative output, raw performance, and policy agendas (see, e.g., [2, 4, 3]), there is a lack of studies with a process-oriented approach to track the concrete procedural steps that laws undergo during their lifecycle. In this paper, we apply to this domain process-oriented techniques from the field of Process Mining (PM). PM is an interdisciplinary research area that aims at extracting insights and knowledge from execution traces of a process, bridging the gap between data science and process science [5].

The Italian Parliament is bicameral, composed of two chambers with equal powers: the *Camera dei Deputati* (Chamber of Deputies) and the *Senato della*

Repubblica (Senate of the Republic). Laws can originate in either chamber and follow a legislative process that requires approval by both. Alongside ordinary laws, the system also includes *decree-laws* (*decreti-legge*), urgent government measures that must be converted into law within 60 days. The procedural history of each law is recorded in unstructured textual documents known as *preparatory works* (it. *lavori preparatori*), official records of the parliamentary events and institutional actions leading to the law's approval. In Italy, these are available starting from 1985 via the Normattiva portal [6] managed by the Presidency of the Council of Ministers. The portal provides a rich but unstructured source of information, covering all laws passed by the Italian Parliament.

We leverage this massive source of information by constructing `ProLiFIC` (*Procedural Lawmaking Flow in Italian Chambers*), a machine-accessible dataset containing Normattiva's preparatory works. These have been first curated, and then encoded as an event log. `ProLiFIC` covers the Italian lawmaking process from 1985 to March 2025, the timeframe available in Normattiva, for overall 5208 laws. By systematically converting the textual content of preparatory works into a structured event log, we offer a detailed representation of legislative activities and procedural steps suitable for automated PM analyses.

Extracting legislative events from natural language, however, poses significant challenges. Preparatory works are written in discursive, often ambiguous, language. As such, manual retrieval and analysis are time-consuming and impractical for large-scale studies. At the same time, rigorous approaches based, e.g., on parsers may be problematic. Preparatory works may, in fact, contain inconsistent formatting over the years, or factual errors (e.g., typos, see Section 4).

To overcome these limitations, we designed a multi-stage pipeline consisting of: (i) web scraping of preparatory texts from Normattiva; (ii) event extraction from the downloaded corpora using large language models (LLMs); and (iii) post-processing and validation to ensure data quality.

The process of Italian lawmaking has been already studied, in part. The authors of [3] use the ILMA dataset [7] to study the temporal dimension of lawmaking. This is a very thorough dataset, with a wider scope than `ProLiFIC`. For example, it also includes information on parliamentary votes, parties, etc. However, these studies use less fine-grained data on procedural aspects, do not use process mining techniques, and consider only a limited time period (1987–2008)[1].

`ProLiFIC` may support rich and diverse types of inquiries in political and legal studies, in part discussed in Sections 5 and 6. These studies complement a growing body of work in European legislative studies, where procedural data is increasingly valued (e.g., [8]). Most notably, `ProLiFIC` allows the application of the rich PM toolbox to legislative data, offering the ability to discover procedural dynamics, compare different lawmaking instances, uncover temporal patterns, and detect inefficiencies in the Italian lawmaking process.

By providing a structured, scalable, and reproducible approach to modeling legislative processes, our work also contributes to the field of legal informatics,

[1] Notably, as of this writing the project appears to be discontinued and unavailable `https://159.149.130.120/ilma/sito/`.

demonstrating how natural language processing (NLP), large language models (LLMs), and data science can enhance the transparency and efficiency of legal systems by introducing *structure* into otherwise unstructured legislative texts [9, 10].

Our aim is therefore dual. First, we present a new dataset. Then, we demonstrate its potential by performing preliminary exploratory data analyses and PM tasks (Sections 5 and 6, resp.). Our analyses focus on the last four Italian legislatures, from 2008 to the present (legislatures XVI to XIX), which we us as a testbed to obtain preliminary empirical insights from real parliamentary data. Exploratory data analysis is done using Python (libraries Pandas and PM4py [11]), while PM analyses with Fluxicon Disco [12]. Thanks to these analyses we can study the following two research questions: **(RQ1)** How do legislative trajectories differ between ordinary laws and the conversion of government decrees? **(RQ2)** How do procedural patterns vary between the two chambers?

2 Related Work and Research Context

2.1 Prior domain-specific datasets

Though no longer publicly accessible, the Italian Law-Making Archive (ILMA) [7] is a fundamental effort in the study of the Italian lawmaking process. Designed for political science research, ILMA provided a relational database that integrated data on laws, legislative initiatives, parliamentary actors, and party positions from the X to the XV legislature (1987–2008).

While our work is inspired by ILMA, it diverges significantly in both scope and methodology. Unlike ILMA, which aimed to provide a comprehensive and centralized overview based on structured institutional metadata, our goal is to construct a fine-grained event log tailored for PM tasks. For this reason, we rely on a different data source (unstructured narrative documents, namely the preparatory works available in Normattiva) and introduce a novel data extraction approach exploiting LLMs (see Section 4). We also broadened the temporal scope, covering a larger and more current view of Italian legislative activity.

2.2 LLMs and NLP for (Legal) Process Mining

In recent years, the PM community has shown growing interest in the intersection with NLP and generative AI, leading to several dedicated initiatives including workshops on NLP for BPM [2]. The need to automatically translate natural language texts into structured representations suitable for business process management (BPM) tools has been recognized for some time. In response, several NLP-based frameworks have been proposed (see, e.g., [13]). Among the most significant contributions in this area, the capabilities of LLMs across a

[2] `sites.google.com/view/nlp4bpm2025` , `www.genai4pm2024.info`

variety of PM tasks, including process querying, model generation, and the interpretation of complex procedural logic have been explored by [14] and further extended in [15], which provides the most recent benchmarks on the subject. These demonstrate that LLMs are highly proficient in understanding process structures, showcasing their potential as general-purpose tools for the domain.

With regard to event log generation specifically, [16] introduces CSV-PM-LLM-Parsing, a Python library that uses LLMs to automatically transform CSV files containing unstructured events into PM event logs. This work shows how LLMs can support data cleaning and standardization.

This literature review highlights a further contribution of our work to this area. We propose a novel use of LLMs in PM: constructing structured event logs from unstructured (and unannotated) legal texts. Rather than using LLMs to clean or standardize already partially structured data (e.g., CSV files), we employ them as semantic extractors capable of identifying and isolating discrete events from raw legislative texts written in natural language, a task that poses significant challenges to legal process discovery [17].

3 On the Italian Legislative Process: Structure, Executive Influence, and European Integration

The Italian lawmaking process reflects an evolving system shaped by both constitutional heritage and modern institutional demands. This process has progressively adapted to the operational needs of a legal framework increasingly influenced by executive exigencies arising from European integration [18].

Central to the Italian legislative model is the notion of perfect bicameralism. As prescribed by Article 70 of the Constitution, legislative authority is exercised collectively by both the Chamber of Deputies and the Senate. Consequently, no bill may become law unless it is approved in identical form by both chambers. This requirement sets in motion a procedural cycle that begins with the submission of a legislative proposal, followed by detailed scrutiny and potential approval in the originating chamber. The text is then passed to the second chamber, which undertakes its own review. It is common for the text to undergo revisions, prompting a back and forth between the two chambers, a process colloquially referred to as the *navetta* (*shuttle*)

Legislative initiative in Italy may originate from multiple institutional actors. Once presented, a bill is assigned to a parliamentary committee, which plays a central and substantive role in the legislative process. It is within these committees that the bulk of the legislative work takes place. Committees are responsible not only for the in-depth examination of the bill, but also for coordinating multiple revisions, consolidating similar proposals, collecting expert opinions, and discussing amendments. After successful passage in one chamber, the bill advances to the other, where a parallel examination occurs. When an identical version is approved by both chambers, the proposal proceed to the President of the Republic for promulgation. The President may return the bill with comments for reconsideration, but once reapproved, promulgation becomes obligatory. The law is then published in the *Gazzetta Ufficiale* (with a unique identifier) and, unless otherwise specified, enters into force 15 days after publication.

In addition to the ordinary legislative process, Article 77 authorizes the issuance of *decreti-legge*, or decree laws, which are temporary legislative acts with immediate legal force. These measures must be ratified by Parliament within 60 days of issuance, or they become null and void retroactively. Originally conceived as instruments for emergency use, decree laws have become a recurring tool in legislative practice. Their widespread use has prompted debate regarding the balance of power between the executive and legislative branches [19]. Nevertheless, decree laws remain a central feature of legislative output and are frequently employed also to expedite politically sensitive measures. This mechanism also contributes to broader structural issues within the Italian legislative system. One such issue is the tendency toward *de facto* unicameralism: the urgency of decree conversion and tight timelines often shift the balance of deliberation in favour of a single chamber, marginalising the role of the other in practice [20].

In addition to the increasing use of decree laws, European integration has also significantly influenced the development of Italy's lawmaking. The country's membership in the EU requires it to transpose directives into national law and to ensure domestic systems are compatible with European regulations and case law [21]. The primary instrument used for transposing directives is the annual *legge di delegazione europea* (European delegation law), which delegates authority to the government to adopt legislative decrees in line with European obligations.

To sum up, the Italian legislative process exhibits a dual character: it is overbureaucratic and procedurally complex, yet increasingly shaped by executive decisions and supranational commitments. The coexistence of multiple legislative routes (ordinary laws, decree laws, and integration of European directives) offers a unique field for legal analysis, which can significantly leverage on process-oriented data science techniques, and in particular process mining ones.

4 Dataset Construction and Description

Here we provide a detailed description of the pipeline to build `ProLiFIC`. We discuss the specific challenges posed by the semi-structured nature of the source data, explaining how these challenges informed our methodological choices (Section 4.1). A concrete example is provided using the case of Law No. 41/1987 (Section 4.2). Finally, we present the structure of the resulting dataset, which consists of two interlinked components: a *metadata table* and an *event log* that captures the procedural trajectory of each law (Section 4.3).

4.1 From Raw Text to Structured Data

`ProLiFIC` has been constructed in three steps: (i) data collection by web scraping Normattiva for preparatory works, (ii) LLM-based event extraction, (iii) post-processing and validation. A visual overview of the pipeline is given in Figure 1.

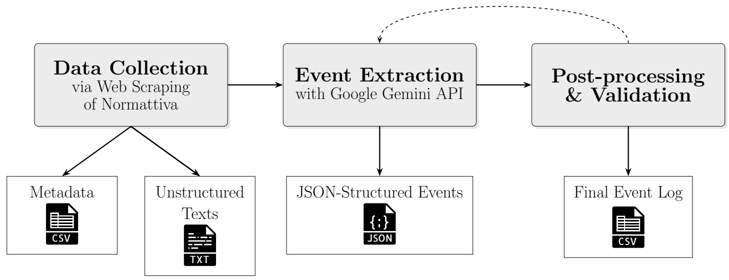


Fig. 1: Dataset construction pipeline.

Data Collection: Scraping Normattiva. As introduced earlier, we began by scraping the Normattiva portal. This provided centralized access to all preparatory works published in the *Gazzetta Ufficiale* since 1985, allowing for comprehensive coverage of the legislative process. For each law, we collected: (i) the full text of the preparatory works, and (ii) a set of metadata fields including the title, description, publication date, and other attributes. Section 4.3 discusses in greater detail the metadata fields. We currently do not make available the dataset, as it is subject to further extensions. We will make it available in an extended version of this work.

Event Extraction via LLMs. The main challenge in transforming the collected data into a processable event log lay in the semi-structured nature of the texts. While some recurring phrasing and formatting conventions exist, they are neither consistent nor sufficient for reliable rule-based parsing. In our initial attempt, we sought to leverage these patterns using traditional parsing techniques such as regular expressions and custom rule-based parsers, supplemented by basic NLP pre-processing steps (e.g., abbreviation expansion, text normalization). However, this method proved inadequate due to the high heterogeneity of the corpus resulting from *ambiguously grouped dates* (e.g., “Esaminato in Aula il 10 giugno 2014; il 3, 4 e 8 luglio 2014 e approvato il 9 luglio 2014.”), *inconsistent labels for commissions and chambers* (e.g., “VIII Comm.” vs. “8ª Commissione”), and *typographical errors* both in texts and dates, to name just a few.

These factors rendered deterministic and exact approaches unreliable, motivating the shift to LLMs for semantic parsing and event extraction. We used the python APIs of Google’s Gemini 2.5 Flash (and Gemini 1.5 Pro for longer texts exceeding the maximum tokens allowed by 2.5 Flash). Thanks to a tailored prompt, we parsed each narrative paragraph into a chronological sequence of standardized procedural events – namely, presentation (*presentazione*), commission assignment (*assegnazione a commissione*), request for opinion (*richiesta di parere*), commission examination (*esame in commissione*), report to the assembly (*relazione all’aula*), plenary discussion (*esame in aula*), and approval (*approvazione*). Extracted event logs were returned in JSON format. Moreover, to facilitate reproducibility and enable future corrections, each extracted event in the dataset was linked to the corresponding excerpt (`chunk`) from the original legal source. This design choice ensures that all procedural inferences remain fully traceable to their textual origin, thereby allowing for manual verification. Given the volume and complexity of the corpus, this is an essential feature.

Post-Processing and Validation. Following the LLM-based extraction, we conducted a rigorous post-processing phase to ensure the accuracy and internal consistency of the event log. We found several errors. Notably, the vast majority of traced errors could not be attributed directly to the LLM's output. Instead, almost all detected anomalies originated from irregularities in the source material, such as typographical errors or formatting inconsistencies in the official legal texts. The post-processing activities included the *normalization of timestamps to minute-level resolution*, *validation of event labels against a fixed taxonomy*, and *correction of start and end markers* using PM techniques (via the pm4py library [11]), to name just a few. We also performed manual reviews to adjust for typographical errors in the original texts, and flagged any procedural trajectories that could not be reconstructed reliably. These records were excluded from the final dataset to preserve overall data quality.

4.2 A case study: Law No. 41/1987

As a concrete illustration of our pipeline, we examine Law No. 41/1987, which established the Sant'Anna School of Advanced Studies. This case demonstrates how complex legislative records can be converted into a structured event log.

Step 1: Raw Preparatory Works. We begin showing the preparatory works associated with the law, which we have crawled from Normattiva. In particular, we show a simplified and translated excerpt from these documents, which serves as the input to our extraction pipeline:

Chamber of Deputies (Bill No. 3780):
Presented by the Minister of Public Education (FALCUCCI) on May 21, 1986.
Assigned to the 8th Commission (Education), in a legislative session, on September 17, 1986, with opinions requested from the 1st and 5th Commissions.
Examined by the 8th Commission on December 11, 1986. Approved on December 18, 1986.

Senate of the Republic (Bill No. 2115):
Assigned to the 7th Commission (Public Education), in a deliberative session, on January 20, 1987, with opinions requested from the 1st and 5th Commissions.
Examined and approved by the 7th Commission on January 29, 1987.

Step 2: Extracted Event Log. Using our LLM-based extraction approach, we then convert the narrative text into a structured event log. We performed a number of iterations of prompt engineering trying to capture the essential procedural events such as presentations, commission assignments, discussions and approvals, along with associated metadata like dates, institutional actors, and commissions involved. For presentation reasons, we omit the actual prompt. The results obtained are shown in Table 1. We can see that each basic activity of the process has indeed been transformed into an actual event in the log.

Date	**Chamber**	**Activity**	**Commission**	**Actor**
21/05/1986	Chamber of Deputies	Presentation	–	Minister Falcucci
17/09/1986	Chamber of Deputies	Assignment	8th Commission	–
17/09/1986	Chamber of Deputies	Request for Opinion	1st, 5th Commissions	–
11/12/1986	Chamber of Deputies	Examination	8th Commission	–
18/12/1986	Chamber of Deputies	Approval	8th Commission	–
20/01/1987	Senate	Assignment	7th Commission	–
20/01/1987	Senate	Request for Opinion	1st, 5th Commissions	–
29/01/1987	Senate	Examination	7th Commission	–
29/01/1987	Senate	Approval	7th Commission	–

Table 1: Extracted event log for Law No. 41/1987. Translated in English for presentation reasons

Step 3: Process Visualization. Finally, we visualize the extracted sequence in a simplified process diagram (Figure 2). Such visualizations can be obtained with virtually any PM tool, like e.g., pm4py or Disco. These graphical representations not only enhance interpretability for human analysts but also serves as a basis for downstream PM tasks such comparative analysis across legislative cases.

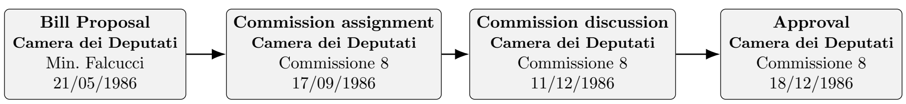


Fig. 2: Simplified process of Law No. 41/1987 establishing Sant'Anna School.

4.3 Dataset Description

The output of our pipeline is a structured dataset that captures both the substantive content (the metadata part of `ProLiFIC`) and the procedural history (the actual event log). It consists of two complementary components that can be linked via a shared identifier, `case_id`, which corresponds to the *codice redazionale* assigned to the law upon publication in the *Gazzetta Ufficiale*. Together, these components offer a rich foundation for both descriptive statistics and process mining analyses of lawmaking in the Italian Parliament, enabling researchers to trace legislative lifecycles, identify bottlenecks, and explore institutional dynamics across changing political configurations.

Metadata part of `ProLiFIC`. Table 2 shows the attributes stored in the metadata. They include general information on each law, such as formal characteristics and dates. Notably, the straightforward structure of this component also allows for future extensions to include additional informative dimensions.

Attribute	Description
`case_id`	Unique identifier for the law
`title`	Official title of the law
`publishing_date`	Date of publication in the *Gazzetta Ufficiale*
`implementation_date`	Date the law takes effect
`decree_conversion`	Indicates if it's a decree-law conversion
`description`	Short summary of the law's purpose
`articles`	Number of articles in the law
`full_text_url`	Link to full legal text on Normattiva

Table 2: Metadata attributes for each law. One row per law.

Event Log part of `ProLiFIC`. Table 3 illustrates the information contained in this component, which captures the procedural trajectory of each law as a sequence of time-stamped events, all sharing the same case id, allowing for detailed reconstruction and analysis of the lawmaking process.

Attribute	Description
`case_id`	Unique identifier for the law
`activity`	Type of legislative action (e.g., proposal, approval)
`time`	Timestamp of the activity, with granularity given in days
`commission`	Commission involved in the event (if applicable)
`person`	Sponsor, speaker, or institutional actor
`chunk`	Extract of legal text or speech related to the activity
`chamber`	*Camera* or *Senato*
`legislature`	Legislative session number (XVII–XIX)
`government`	Executive in office at the time

Table 3: Event log with all events of all laws. One row per event.

5 Exploratory Data Analysis on Recent Legislatures

Here we shed further light on `ProLiFIC` by providing a picture of the data contained in four reference legislatures. In particular, we focus only on the most recent ones, namely XVI (2008-2013), XVII (2013–2018), XVIII (2018–2022), and XIX (2022–March 2025), for a total of 1244 laws (both decree-laws and ordinary laws). We remark that, at the time of writing, legislature XIX has not concluded yet. Therefore, we have data for around 50% of its overall length.

Legislature	Laws *(total)*	% Decrees	Median Duration *(days)*	
			Ordinary	Decrees
XVI	373	30%	159	53
XVII	367	23%	375	54
XVIII	312	33%	413	54
XIX *(ongoing)*	192	38%	251	53

Table 4: Legislative activity: counts and median duration by legislature. Median duration is shown separately for ordinary and decree laws.

Table 4 shows, for each legislature, the total number of enacted laws (including the percentage of decree-laws) and the median duration of the legislative process, calculated as the number of days elapsed between the first and last recorded event in the law's procedural history.

These numbers reveal some meaningful patterns. While the total number of laws passed remains relatively stable across the four legislatures, their durations vary considerably. In particular, Legislature XVIII stands out for ordinary laws: while enacting a slightly lower, yet comparable number of laws, it shows a notably higher median duration for ordinary laws (413 days), nearly double that of Legislatures XVI and XIX. This pronounced delay raises important questions about potential procedural or political factors that may have contributed to the slowdown during that term, possibly related to the COVID-19 pandemic and the instability of parliamentary majorities.

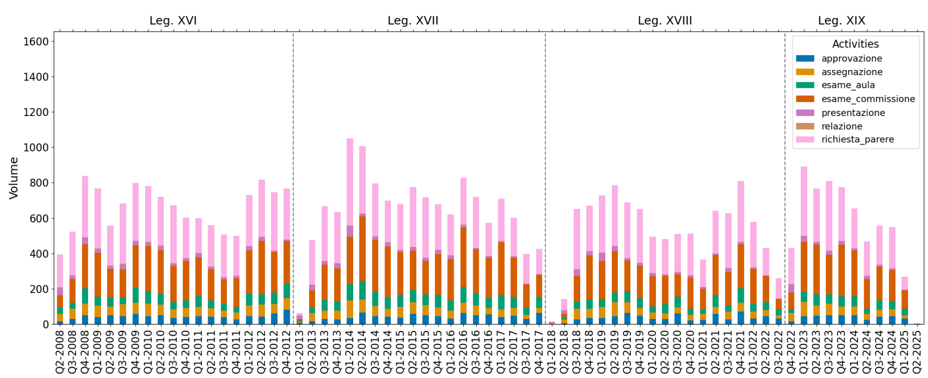


Fig. 3: Frequencies of events. Dashed lines separate the different legislatures.

This is further illustrated from a different perspective in Figure 3. We analyze a disaggregated view of parliamentary activity, focusing on the volume of activities indicating "hops" between committees and chambers. The figure shows strong fluctuations in legislative intensity over time. While a comprehensive explanation of these trends lies beyond the scope of this paper, we highlight a few notable examples. The longer median durations observed during the COVID-19 period appear to coincide with a marked decline in legislative output. However, this may not be the primary factor explaining the most visible peaks. Similar slowdowns in other legislatures suggest that the heterogeneity of governing majorities and the fragmentation of the parties might offer another plausible explanation. E.g., a sharp increase is observed between the fourth quarter of 2013 and the first one of 2014, coinciding with the establishment of a strong and cohesive government majority. Conversely, the first two quarters of 2018 show a significant drop, reflecting a period of political instability.

Returning to Table 4, some notable insights emerge regarding decree-laws. First, there is a clear trend of increasing reliance on this legislative instrument, as evidenced by the growing percentage of decree-laws over time. Second, and perhaps more striking, is the remarkable consistency in their median duration across legislatures. In contrast to ordinary laws, the duration of decree-laws remains in fact consistent across legislatures. However, this is not surprising, as decree laws must be ratified by Parliament within 60 days of issuance, otherwise they become null. This reinforces the notion that different legislative issues follow distinct procedural path and highlights the value of examining not just how many laws are passed, but how they are processed. In the next section, we turn to PM techniques to investigate these procedural dynamics in greater depth, analyzing the structure and frequency of legislative activities.

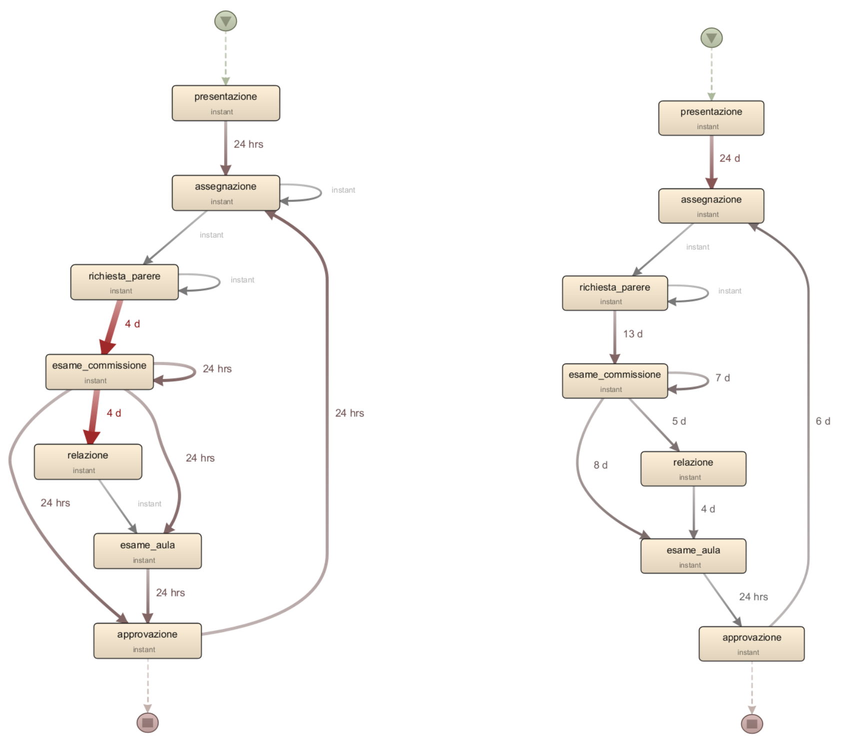
6 Process Mining with Fluxicon Disco

To demonstrate the analytical potential of structuring legislative data as event logs, we used *Fluxicon Disco* [12], a leading PM tool. We reconstruct and compare the procedural paths of two major categories of legislation: ordinary legislative proposals and decree conversion laws.

Category	**Variants**	**Median Duration**
Ordinary laws	579	9.6 months
Decree conversions	370	54 days

Table 5: Comparison of ordinary laws and decree conversions by number.

Disco allows us to identify variants, i.e., process instances with the same events (ignoring timings), to measure durations, and to detect structural patterns. This is reported in Table 5. We see that ordinary laws follow a greater number of variants and exhibit significantly longer durations compared to decree conversions. This is consistent with what we expected, i.e., ordinary laws involve more iterative steps, suggesting a less streamlined and more deliberative process, while conversions of decrees are prioritised (indeed, they are required to be approved by Parliament within 60 days). Most interestingly, although the timelines differ significantly, the processes remain largely the same: the same activities are carried out, but at completely different paces. This is shown in the process maps obtained with Disco shown in Figure 4. Disco allows setting *importance filters* to focus on more or less frequent/important activities (the boxes) and paths (actually, edges). In order to preserve all activities while focusing on only the most important relations, we have set 100% and 0% for activities and paths, respectively.



(a) Maps for decree conversion laws

(b) Map for ordinary laws

Fig. 4: Disco's process maps showing median durations. We have set 100% and 0% as importance zoom for activities and paths, respectively.

Furthermore, we can visualize the differences in the work of the two chambers with the diagram in Figure 5, where we have refined the first activity of laws (presentazione) by denoting whether it has been initiated by the Senate (Senato) or by the Chamber of Deputies (Camera). Then, using Disco's filters, we filtered for laws initiated in one of the two chambers only (not shown in the figure). We find out that there is no relevant divergence in the process of laws initiated in either chamber. Instead, it is important to note that, as shown in Figure 5, the Senate is consistently more expedite than the Chamber in every step of the *iter*

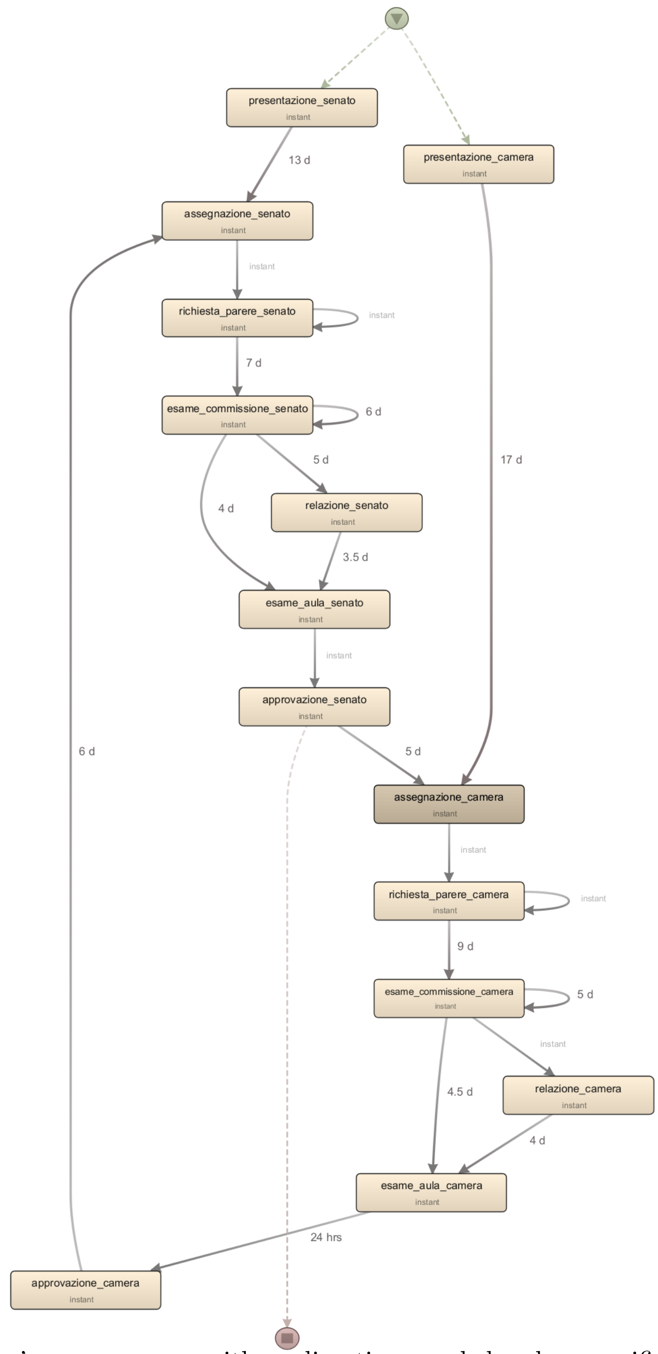


Fig. 5: Disco's process maps with median times and chamber-specific events. We have set 100% and 0% as importance zoom for activities and paths, respectively.

legis independently of the starting chamber. Studying the reasons of these asymmetries is beyond the scope of this work. Here we remark that, despite under perfect bicameralism the two organs have the same functions and prerogatives, they do differ in: Internal Rules of Procedures; Active and passive electorate (minimum 25 and 40 years old to be elected in the Chamber and Senate, respectively – until 2021, minimum 25 years old to vote for the Senate, now uniformed to the minimum voting age of 18); Numbers, 200 Senators (315 before 2020) and 400 Honourables (630 before 2020). Further analyses on such divergence could be carried out using `ProLiFIC`. We plan to do it in future works.

7 Conclusions

We presented `ProLiFIC` (*Procedural Lawmaking Flow in Italian Chambers*), an event log of preparatory works of Italian laws, together with preliminary exploratory data and Process Mining analyses. We built and applied an LLM-based pipeline to unstructured texts from the Normattiva portal, extracting structured event logs capturing the Italian lawmaking process for 1985-2025. We discussed the methodological challenges in the creation and curation of the dataset.

The preliminary analyses allowed us to address two research questions of interest to the Law and Political Science: **(RQ1)** *How does the legislative process differ between ordinary laws and decree conversions?* and **(RQ2)** *How do procedural trajectories vary between the Chamber of Deputies and the Senate?* Although preliminary, our findings reveal meaningful patterns in both cases, demonstrating that this type of data can support substantive empirical research.

Beyond these initial insights, the dataset establishes a foundation for more comprehensive investigations into legislative behavior, institutional dynamics, and the temporal structure of lawmaking.

Future work. On the dataset side, we plan to extend its coverage and enhance its richness by incorporating additional dimensions from complementary sources. E.g., we aim to explore the feasibility of linking each law to its corresponding policy domain, enabling domain-specific analyses. We will also improve and systematize the validation of the dataset. A revised version of the dataset is presented here [22]. We may also consider a wider domain, e.g., European Laws. On the analysis side, we plan to extend our PM analyses using conformance checking techniques (e.g., [23]), moving beyond the visual exploration enabled by Disco's maps. To support this, we will integrate the PM4Py library into our analytical workflow. More in general, we will deepen the presented analyses, e.g., by explaining the reasons behind the discovered asymmetries in the performance of the two chambers, and by further analysing the impact of COVID.

Acknowledgments. The work has been partially supported by project SMaRT COnSTRUCT (CUP J53C24001460006), in the context of FAIR (PE0000013, CUP B53C22003630006) under the National Recovery and Resilience Plan (Mission 4, Component 2, Line of Investment 1.3) funded by EU - NextGenerationEU.

References

1. Capano, G., Giuliani, M.: Governing without surviving? an italian paradox: Lawmaking in italy, 1987-2001. The Journal of Legislative Studies **7**(4) (2001) 13–36
2. Kreppel, A.: The impact of parties in government on legislative output in italy. European Journal of Political Research **31**(3) (1997) 327–349
3. Borghetto, E., Giuliani, M.: A long way to tipperary: Time in the italian legislative process 1987–2008. South European Society and Politics **17**(1) (2012) 23–44
4. Meneghetti, N., Pacini, F., Biondi Dal Monte, F., Cracchiolo, M., Rossi, E., Mazzoni, A., Micera, S.: Predicting party switching through machine learning and open data. iScience **26**(7) (2023) 107098
5. van der Aalst, W.: Process Mining: Data Science in Action. (2016)
6. Presidenza del Consiglio dei Ministri: Normattiva. `https://www.normattiva.it/`
7. Borghetto, E., Curini, L., Giuliani, M., Pellegata, A., Zucchini, F.: Italian lawmaking archive (ilma): A new tool for the analysis of the italian legislative process. Rivista Italiana di Scienza Politica **3** (2012)
8. Reh, C., Héritier, A., Bressanelli, E., Koop, C.: The informal politics of legislation: Explaining secluded decision making in the european union. Comparative Political Studies **46**(9) (2013) 1112–1142
9. Caponecchia, V., D'Agostino, B., Comandè, G., Licari, D., Vandin, A.: Towards visualizing and analysing legal proceedings with process mining. In: Proceedings of PLC 2024, co-located with ICPM 2024. Volume 3850 of CEUR. (2024) 46–57
10. Caponecchia, V., D'Agostino, B.S.A.S., Comandè, G., Licari, D., Vandin, A.: Process mining for legal courts: Visualising, analysing and comparing italian divorce proceedings. Computer Law & Security Review (2025)
11. Berti, A., van Zelst, S.J., van der Aalst, W.M.P.: Process mining for python (pm4py): Bridging the gap between process- and data science. CoRR (2019)
12. Günther, C.W., Rozinat, A.: Disco: discover your processes. In: BPM Demo. (2012)
13. Sànchez-Ferreres, J., Burattin, A., Carmona, J., Montali, M., Padró, L., Quishpi Betun, L.: Unleashing textual descriptions of business processes. Software and Systems Modeling **20** (2021)
14. Berti, A., Schuster, D., van der Aalst, W.M.P.: Abstractions, scenarios, and prompt definitions for process mining with llms: A case study (2023)
15. Berti, A., Kourani, H., van der Aalst, W.M.P.: PM-LLM-Benchmark: Evaluating large language models on process mining tasks. In: PM Workshops. (2025)
16. Berti, A., van der Aalst, W.M.P.: CSV-PM-LLM-Parsing: Automatic ingestion of CSV event logs for process mining using llms. In: BPM Demo 2024
17. López, H.: Challenges in legal process discovery. In: ITBPM, CEUR (2021) 68–73
18. Bartole, S., Bin, R., Pitruzzella, G.: Diritto costituzionale: profili teorici ed evoluzione giurisprudenziale. Giappichelli, Torino (2022)
19. Redi, C.: Il decreto-legge: strumento di legislazione "straordinariamente ordinaria". In: Osservatorio sulle fonti – Speciale Decreto-Legge. (2014)
20. Lolli, I., et al.: Decreti-legge e disegni di legge: il governo e la'sua'maggioranza. Osservatorio sulle fonti (3) (2016) 1–59
21. Morosini, M.: L'attuazione del diritto dell'eu nel più recente periodo: legge di delegazione europea e legge europea alla luce della prassi applicativa. Osservatorio sulle fonti – The Implementation of EU Law in Member States **2** (2017)
22. Contestabile, M., Ferrara, C., Giovannetti, A., Parrillo, G., Vandin, A.: The ProLiFIC dataset: Leveraging LLMs to Unveil the Italian Lawmaking Process (2025)
23. Incerto, E., Vandin, A., Ahrabi, S.S.: Stochastic conformance checking based on variable-length markov chains. Inf. Syst. **133** (2025) 102561