Standardization and harmonization of law through automated process analysis and similarity techniques

Davide Audrito[1], Luigi Di Caro[2], Laura Genga[3], Rachele Mignone[2], Roberto Nai[2,*], Ivan Spada[2], Emilio Sulis[2] and Vittoria Margherita Sofia Trifiletti[4]

[1]*Department of Legal Studies - University of Bologna, Via Zamboni 27/29, 40126, Bologna, Italy*
[2]*Computer Science Department - University of Turin, C.so Svizzera 185, 10149, Torino, Italy*
[3]*School of Industrial Engineering - Eindhoven University of Technology, Eindhoven, The Netherlands*
[4]*Department of Law - University of Turin, C.so Lungo Dora Siena 100, 10153, Torino, Italy*

Abstract

In comparative law, standardization and harmonization of law are interrelated concepts that aim to increase legal certainty, efficiency, and accessibility. Standardization focuses on creating uniform rules and procedures within a specific jurisdiction, while harmonization seeks to align legal frameworks across jurisdictions. This research aims to investigate the impact of standardization and harmonization on legal systems through the application of AI methods. Two case studies were analyzed using machine learning, natural language processing, and process mining techniques. The paper also reports a brief discussion of the methods adopted and early results from two legal domain experts.

Keywords

Process Mining, Discovery of legal processes, Harmonization of Law, Legal Standardization

1. Introduction

Comparative law, by examining the similarities and differences between various legal systems, plays a crucial role in promoting the harmonization and standardization of legal processes across different jurisdictions [1]. These topics are particularly relevant in the legal system of the European Union (EU), where the aim is to create a more consistent and unified legal framework among member states. As a matter of fact, both the harmonization of laws and public procurement are crucial processes for fostering fair competition and non-discrimination among European member states.

In recent years, legal decision making has benefited from analyzing large volumes of data using computational tools, which allow knowledge to be extracted automatically [2]. In legal informatics, increasing attention has been given to automated methods of extracting knowledge from legal sources [3]. The research on artificial intelligence (AI) includes techniques and operational tools for legal text analysis using methods from the disciplines of natural language processing (NLP) and machine learning (ML) [4]. In addition, digital traces left in information systems (event logs) have been analyzed with algorithms and tools that can reconstruct and investigate the flow of events, in the context of process-oriented data science. The recent research in Business Process Management (BPM) offers a view of how processes are actually executed by revealing hidden patterns, inefficiencies, and opportunities for improvement [5]. The process model can be reconstructed from the data and also simulated [6]. The most recent discipline of process mining concerns the automatic analysis of processes, and originates from BPMN but can also be applied in other domains, e.g. healthcare [7] or education [8, 9].

PLC - Processes, Laws and Compliance workshop, in conjunction with ICPM 2024, October 14, 2024, Lyngby, Denmark
*Corresponding author.
d.audrito@unito.it (D. Audrito); luigi.dicaro@unito.it (L. D. Caro); l.genga@tue.nl (L. Genga); rachele.mignone@unito.it (R. Mignone); roberto.nai@unito.it (R. Nai); ivan.spada@unito.it (I. Spada); emilio.sulis@unito.it (E. Sulis); vittoriamargheritasofia.trifiletti@unito.it (V. M. S. Trifiletti)
0000-0002-0877-7063 (D. Audrito); 0000-0002-7570-637X (L. D. Caro); 0000-0002-9421-8566 (L. Genga); 0009-0009-2699-8730 (R. Mignone); 0000-0003-4031-5376 (R. Nai); 0009-0002-0459-1189 (I. Spada); 0000-0003-1746-3733 (E. Sulis); NA (V. M. S. Trifiletti)

This article summarizes and discusses some recent research contributions in the practical case of automated analysis of legal harmonization and public procurement law in the European Union (EU). The methods we present in the article and the initial results obtained from ongoing research are encouraging towards further adoption of advanced information technology tools in comparative law. Indeed, knowledge management and process mining (PM) tools make it possible to facilitate comparative analysis to identify similarities, differences, and possible convergences between different legislation. Finally, two legal experts provided valuable support at all stages of the research, playing a key role in analyzing both methods and results.

The article is structured as follows. Section 2 details the background of the harmonization of European legislation and public procurement studies. Section 3 details the methodologies applied in the analysis, finally proposing a brief review of the initial results obtained in Section 4. Finally, Section 6 concludes the articles with future work.

2. Background

2.1. European legal context

Comparative law involves the systematic comparison of different legal systems to gain a deeper understanding of the law's functions by analyzing the similarities and differences between various legal systems. The discipline plays a crucial role in fostering international legal harmonization, as well as in understanding the diverse legal frameworks governing the acquisition of goods, services, and works by public entities across different countries. In the past decades the use of quantitative methods aside the more traditional qualitative methods of analysis have been very much debated in comparative law literature [10]. Legal harmonization refers to the process of aligning laws and regulations across different jurisdictions to create a more uniform legal framework. Public procurement law governs how public authorities purchase goods and services. The legal standardization of public procurement processes enhances transparency, efficiency, and compliance by establishing uniform rules and procedures, reducing complexity and disputes, and facilitating cross-border participation and fair competition.

In recent years, the integration of computational tools and applications has opened new avenues for comparative law research, enabling more efficient and comprehensive analyses [11, 12]. Information system databases collect valuable data and information that can be analysed with computer techniques and algorithms to extract knowledge. The data used in legal process analysis can be extensive and scattered, particularly for activities like calls for procurement. Online repositories that collect information on European-level procurement offer a rich dataset for researchers to explore.

2.2. Process Mining for law

PM offers a data-driven approach to analyze business processes [13]. Data with timed events from information systems can be transformed into event logs to capture the sequence of activities within a system. Dataset can be merged to obtain more events and information on a process. This is the case, for instance, of a French dataset on public procurement described in [14], which combines dataset at the EU and national level. Event logs enable to uncover the actual processes executed (process discovery) or the different paths (variant analysis), as well as to compare the discovered processes against predefined models to ensure compliance and identify deviations (conformance checking). Such analysis allows the identification of bottlenecks, inefficiencies, and opportunities for automation to enhance overall operational efficiency.

To improve understanding of legal processes, by applying PM techniques, organizations can gain a deeper understanding of their operations, enabling them to make data-driven decisions and improve compliance. Legal issues have so far not been much explored through PM techniques, perhaps because of the difficulty of finding information systems with timed legal events. Nevertheless, some studies investigated legal process discovery to analyze the judicial performance of a Brazilian court [15], the public procurement processes in Croatia [16], the automated analysis of Italian [17] or European public

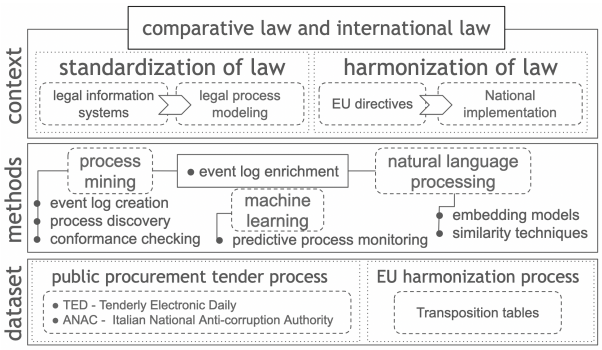

Figure 1: An overview of context, methods, and dataset used in the research on standardization and harmonization of law. Full-size image available at https://github.com/roberto-nai/PLC2024.

procurement processes [18]. Process oriented approaches enable the identification and mitigation of compliance risks by providing a detailed view of actual process execution and deviations from standard operating procedures [19]. The use of PM for audit and assurance provides auditors and compliance officers with a robust tool to assess the effectiveness of internal controls and identify areas where process improvements can enhance compliance [20]. PM can help organizations to demonstrate GDPR compliance by providing visibility into data processing activities and identifying potential data breaches [21].

2.3. Knowledge Management for law

Knowledge management in the legal field has been developed in several research directions. A first area of research focused on the representation and reasoning of legal knowledge in a structured way, such as through deontic logic, rule-based systems, and legal argumentation [22]. Another relevant research area focused on the development of advanced systems for organizing, storing, and retrieving legal documents including the use of semantic search technologies and AI to improve access and accuracy of legal information [23]. A pipeline for exploit the domain expert modus operandi in similarity case matching has been introduced by [24]. Semantic web and legal ontologies help to structure and link legal concepts for allowing machines to understand and reason with legal data [25] and to enhance integration and interoperability of legal systems [26]. This kind of research supports applications like automated legal advice and intelligent legal research [27].

Recent research in the fields focused on AI, ML, and NLP applications in legal practice to automate legal tasks such as document analysis or fraud detection[28], legal advice [27], legal outcome predictions [29, 30] and explain them [31].

3. Methods and techniques

This section outlines the methodologies employed in standardization and harmonization research, introducing two case studies and their respective datasets, as illustrated in Figure 1. The following two sections provide an overview of research techniques used in comparative law and international law, supported by two practical applications.

3.1. Process mining on public procurement process

As stated above, public procurement is a key area within legal standardization, as it involves creating and enforcing uniform legal frameworks that govern how public contracts are awarded and managed. In EU, this process facilitates consistent practices across different national countries and jurisdictions. This section describes the practical application of PM for examining public procurement processes. Figure 2 summarises the approach, starting with the collection of timed data (1st step) from legal

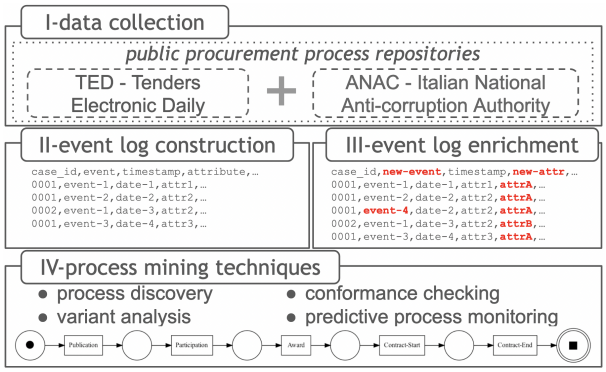

Figure 2: The methodological approach for studying public procurement process with a process-oriented perspective. Full-size image available at https://github.com/roberto-nai/PLC2024.

public repositories. In particular, we refer to a dataset at the EU level, i.e. Tender Electronic Daily - TED[1], merged with data from a National repository on public procurement, i.e. ANAC[2]. The public procurement dataset from 2016 to 2023 covers 3 million tenders is available online and has been described in [32]. TED data includes five events (publication, participation, award, contract-start, contract-end), from the creation of the procurement up to the end of the contract. Italian dataset adds some events in each procurement process, e.g. the contracting company, the presence of variants during the process, or the state of progress of the procurement. The next step is the construction of the event log (2nd step), with information on each case (case identifier), the name of the event or the activity, and the corresponding timestamp. Eventually, some attributes can be added at the level of case or trace. Next, the enrichment of the event log (3rd step) can occur in several directions: new events with new dates as well as new attributes can be extracted from the procurement text (at the case or trace level). This task can be performed with both traditional techniques, such as regular expressions, as well as modern ones such as large language models [27]. PM techniques (4th step) are applied to the final event log, enabling the analysis of the legal process control-flow with process discovery and variant analysis, as well as the automatic analysis of compliance against an ideal model or the application of AI methods such as forecasting.

3.2. National implementing measures and harmonization

While the comparison between reference legislation (e.g., EU directives) and national legislation requiring alignment necessitates substantial legal expertise, this study endeavours to explore the feasibility of incorporating automated techniques into this process. In this work, we describe a pipeline for applying textual similarity metrics to verify the correspondence of content between European and national legislation, which represents a preliminary level of legal harmonization in Europe.

This NLP pipeline has been explored on a dataset of an EU-funded project[3] that is developing a platform designed as a legal comparison system. This system aims to quantify the extent to which an EU member state has transposed a certain segment of European legislation into its legislation. The process starts with a pre-processing phase, during which European and national legislative segments are subjected to cleaning and normalization to ensure consistent text formatting. This step is followed by the generation of semantic paragraph and word embeddings, both using the *nlpaueb/bert-base-uncased-eurlex* model [33]. Finally, a harmonization metric is computed between the European legislative segments and their corresponding national transpositions, quantifying the alignment level within the dataset. The Harmonization Index (HI) is a metric designed to assess the degree of semantic overlap between two legal texts. It serves as a similarity measure, identifying and emphasizing common elements to compare different legal texts, considering both *symmetrical* and *asymmetrical* relationships. The HI

[1] https://ted.europa.eu/en
[2] https://dati.anticorruzione.it/opendata
[3] The FACILEX project web site: https://site.unibo.it/facilex/en/the-project

was introduced to measure the degree of coherence or uniformity between legal texts across different jurisdictions or entities. This index is particularly valuable in the legal domain because it provides a quantitative means to assess the alignment between national laws and supranational regulations, such as those adopted at the European level, or between local laws and federal regulations [34]. In the context of legal texts, the HI plays a critical role in efforts to harmonise laws, especially in areas such as European law or international law, where there is a continuous drive to align the legal frameworks of various countries. This is particularly relevant in public procurement and regulatory compliance processes, where harmonisation can ensure smoother cross-border transactions and reduce legal uncertainties [35]. The index also serves as an empirical tool to monitor the progress and effectiveness of harmonisation efforts by providing an indication of the level of legal alignment across jurisdictions [36].

This research leverages the HI to gauge the extent to which a European legal act, such as the EU Council Framework Decision on the European Arrest Warrant, has been implemented at the national level. The computation of the HI for a given EU text and its corresponding national text is a complex task as each section of an EU legal act can be transposed into national law through a varying number of legal sources, ranging from zero to potentially numerous sources. The HI is computed for each European legislative segment in two ways: symmetric and asymmetric. The symmetric version, computed at the paragraph level, represents the average *cosine similarity*[4] between the European segment and each of its national transpositions. Conversely, the asymmetric version considers the highest cosine similarity between each word in the European segment and all words within the corresponding national transposition. This version incorporates a weight function applied to each word, reflecting its relative importance in the overall computation. Consequently, the HI is a cumulative measure that accounts for all instances of transposition for each part of the EU text. Following are the symmetric and asymmetric cosine similarity formulas; in the symmetric form, the denominator considers the magnitudes of both vectors, while in the asymmetric form, only the magnitude of vector A is considered.

$$\text{Cosine Similarity (Symmetric)} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

$$\text{Cosine Similarity (Asymmetric)} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|^2}$$

4. Initial results and discussion

This section describes the first results with a brief discussion. A description of the initial dataset opens each presentation. Finally, a brief summary of the challenges collected by the domain experts involved in the analyses closes the section. Note that the initial TED dataset is publicly available[5], while the graphs obtained from the event log come from Fluxicon's DISCO tool[6].

4.1. European public procurement notices

TED Dataset. The extraction of data from the TED web repositories[7] allows to reach 27,252 records for Italian cases in the years 2016 - 2022[8]. For the same years, 99,915 French and 47,743 Spanish cases were also extracted; according to domain experts, the three member states have similar legal review systems.

[4]Cosine similarity measures the similarity between two non-zero vectors by calculating the cosine of the angle between them, ranging from -1 (completely dissimilar) to 1 (completely similar). It can be symmetric, where the similarity between vectors A and B is the same as between B and A, or asymmetric, where the order of the vectors affects the result.

[5]https://github.com/roberto-nai/PLC2024

[6]https://fluxicon.com/disco

[7]https://data.europa.eu/data/datasets/ted-csv?locale=en

[8]The time frame is fairly consistent with the validity of the Italian public procurement code, which came into force in April 2016 and was replaced on 31 March 2023

```
Case ID;Activity;Timestamp;Sector;Amount;Nuts;Country
...
2017106814;PUBLICATION;2017-03-17;S;1035000.0;ITC13;IT
2017106814;PARTICIPATION;2017-05-09;S;1035000.0;ITC13;IT
2017106814;AWARD;2017-06-07;S;1035000.0;ITC13;IT
2017106814;CONTRACT-START;2017-09-01;S;1035000.0;ITC13;IT
2017106814;CONTRACT-END;2022-07-31;S;1035000.0;ITC13;IT
2017107959;PUBLICATION;2017-03-20;S;637622.4;ITE19;IT
2017107959;PARTICIPATION;2017-05-03;S;637622.4;ITE19;IT
2017107959;CONTRACT-START;2017-06-01;S;637622.4;ITE19;IT
2017107959;AWARD;2017-06-15;S;637622.4;ITE19;IT
2017107959;CONTRACT-END;2020-05-31;S;637622.4;ITE19;IT
...
```

Figure 3: Legal event log example in CSV format obtained from TED dataset. Full-size image available at https://github.com/roberto-nai/PLC2024.

Event log construction. The event log is obtained by extracting from the TED dataset the relevant data needed to construct the traces of each procurement. Figure 3 contains an extract of the event log obtained; in addition to the basic elements (Case ID, Activity and Timestamp), the traces also contain the attributes Sector ("Works", "Services", "Supplies"), Amount, NUTS[9] and Country.

Event log enrichment. Event log enrichment with respect to dates can use traditional techniques, such as regex, as well as large language models. In our case study, new events captured from the text files connected to every procurement is the *bid-opening*; in TED, it involves the formal process where received bids are unsealed and examined after the submission deadline.

Further enrichment with national dataset, such as ANAC for example, allows for discover more detailed insights into the process beyond the creation, award, start/end of contract phases; through the ANAC dataset, it is possible to add other procurement events, such as the presence of a variant (i.e., a modification of the initial process), the verification of the status of works, any sub-contracts with other economic operators.

Process mining techniques. Figure 4 illustrates an example of the processes uncovered by the data, which made it possible to immediately identify the similarities and differences in the three process diagrams for each country. The diagrams illustrate the average duration between two activities, indicated by the thickness of the arcs, which also helps identify bottlenecks easily. The average time from the procurement notice to participation is approximately 36 days (minimum 34, maximum 39), while the average time to determine the procurement winner is around 45 days (minimum 36, maximum 54). The most prolonged part of the process is the decision-making phase for the winning procurement, which takes about 19.3 months. In addition to the timing, loops can be observed in the Spanish process, particularly in the PUBLICATION and PARTICIPATION activities. According to domain experts, this occurs because contracting authorities may re-publish a procurement notice, prompting participation to take place again based on the new publication. Another observation is that the AWARD activity does not always occur; according to domain experts, this is due to 'framework agreements' (multi-year contracts across several procurements that are published but automatically awarded).

Figure 5 presents, on the other hand, the newly identified BID-OPENING event extracted from the procurement notices text. This event demonstrates that the opening of received bids occurs promptly (transition from PARTICIPATION to BID-OPENING). At the same time, most of the time required to award the contract to the economic operator is consumed by the decision-making process of the contracting authorities (transition from BID-OPENING to AWARD).

Finally, Figure 6 illustrates how the TED event log was extended with further activities using the ANAC dataset, which contains a specialisation of the TED data with more activities like VARIANTS, SUBCONTRACTS, etc.

[9] https://www.europarl.europa.eu/factsheets/en/sheet/99/nomenclatura-comune-delle-unita-territoriali-statistiche-nuts-

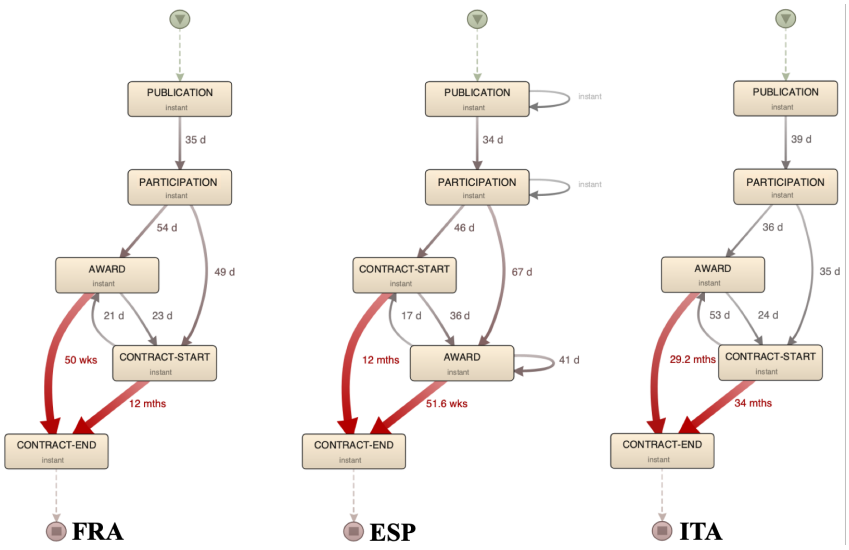

Figure 4: The European public procurement processes of France, Spain, and Italy. Full-size image available at https://github.com/roberto-nai/PLC2024.

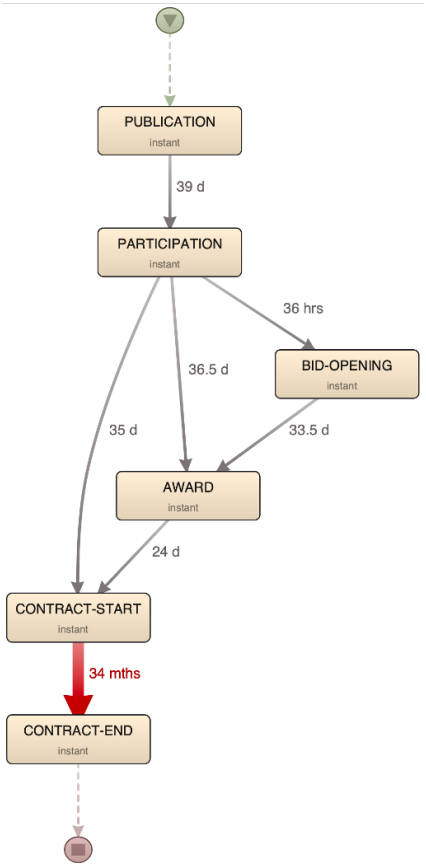

Figure 5: The Italian public procurement process from TED dataset, with the new event BID-OPENING extract from the texts. Full-size image available at https://github.com/roberto-nai/PLC2024.

4.2. Transposition tables to assess legislative harmonization.

TT Dataset. The dataset comprises Transposition Tables (TT) aligning EU legal acts, namely Decision 2002/584, Directive 2014/41, and Regulation 2018/1805, and their corresponding national transpositions as annotated by legal experts. The project focused on French, German, Italian, Portuguese, and Spanish national implementations for each of the above-mentioned European documents.

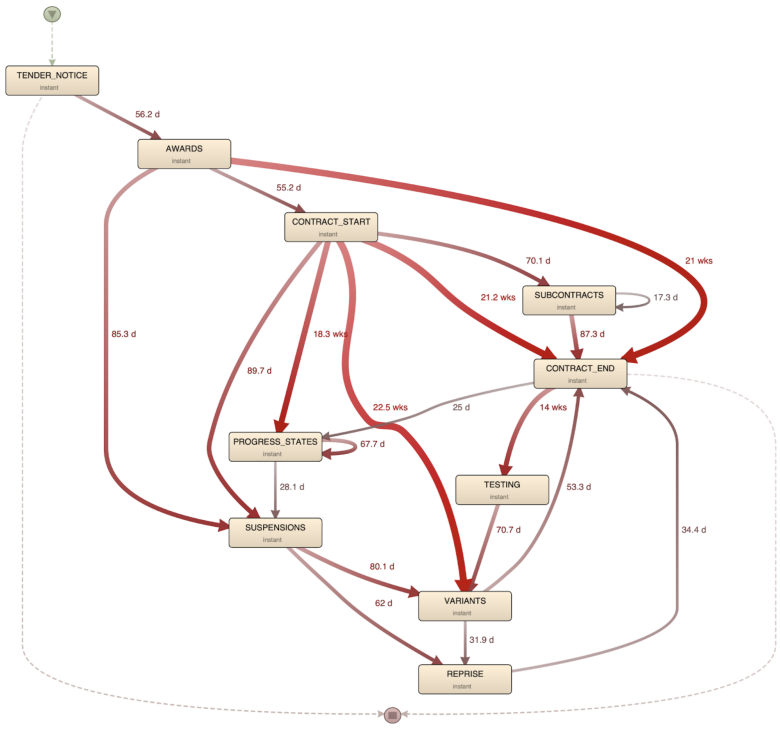

Figure 6: The Italian public procurement process in detail, merging TED and ANAC datasets for "Works" category. Full-size image available at https://github.com/roberto-nai/PLC2024.

This dataset is structured into documents, articles, and paragraphs, each of which may or may not have a corresponding transposition, which is in turn classified as either explicit or implicit. This classification aligns with different levels of national implementation, which include full implementation, partial implementation, and instances where no implementation has occurred at all.

Despite its multilingual nature, the dataset is fully translated into English and available on the project platform, in both tabular and JSON formats, allowing for easier computational analysis.

Harmonization Index results. Segment-level scores were computed following both a symmetrical and asymmetrical approach, resulting in two different distributions (Figures 7 and 8). The symmetrical version exhibited a mean score of 0.93 and a median of 0.95, while the asymmetrical version yielded a mean of 0.73 and a median of 0.69.

Manual analysis indicates that the symmetrical version of the Harmonization Index aligns more closely with the human judgment of legislative transposition completeness than its asymmetrical counterpart. This suggests the symmetrical version is preferable for evaluating transposition levels. However, this finding prompts further investigation into how the index's performance compares to alternative approaches operating at different textual granularities.

5. Considerations of legal experts

PM offers a granular view of how legal processes are actually executed. By analyzing event logs, legal experts appreciated the possibility to reveal hidden patterns, inefficiencies, and opportunities for improvement. With respect to legal standardization, the benefits offered by automated analysis mainly concern the identification of legal process variants and compliance issues. By comparing the actual legal process execution to the ideal process model, PM can easily detect where and why deviations occur. At the same time, inconsistencies can be detected, e.g. areas where different jurisdictions apply laws differently, as the procurement process case study demonstrated.

The methodology enables country-based comparative analyses across different time-frames, although

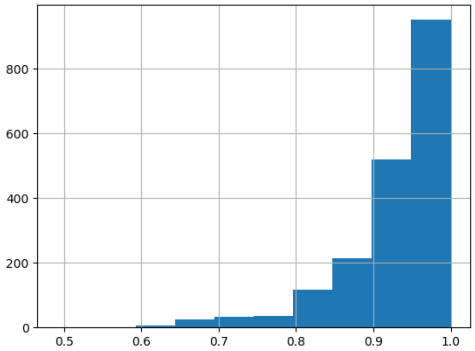

Figure 7: Distribution of similarity scores using the *symmetrical* HI. The x-axis represents the similarity scores, while the y-axis indicates the corresponding count.
Full-size image available at https://github.com/roberto-nai/PLC2024.

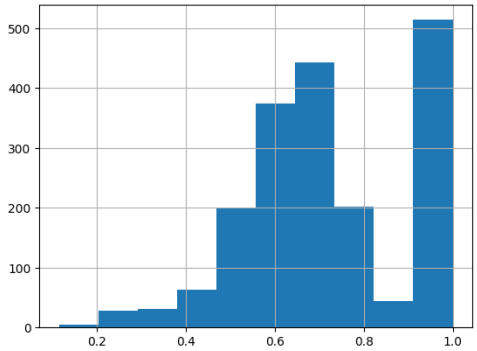

Figure 8: Distribution of similarity scores using the *asymmetrical* HI. The x-axis represents the similarity scores, while the y-axis indicates the corresponding count.
Full-size image available at https://github.com/roberto-nai/PLC2024.

data availability varies significantly across countries on platforms like TED. This disparity can be attributed to differences in resources and priorities in data sharing among nations. The analysis revealed that procurement procedures cannot be reduced to mere legislative frameworks - they require both quantitative and qualitative approaches involving key stakeholders.

In the Italian context, PM has proven particularly valuable for monitoring compliance with recent amendments to the Public Contracts Code. The analysis showed that for services, the mean time between contract award and start is 36.4 days, for works 55.2 days, and for supplies 24.2 days. These timelines can be compared against regulatory requirements, such as the standstill period of 35 days established by Legislative Decree 50/2016, with exceptions for sub-threshold services and supplies.

The process analysis identified several possible bottlenecks, particularly between publication and award phases (approximately 80 days) and between award and contract start (22.5 days). The data revealed that contract duration varies significantly by type - works averaging 31 months, supplies 9.1 months, and services 5.3 months. These differences reflect the distinct nature and regulatory requirements of each category.

Comparative analysis between countries showed that German procedures tend to be longer than Italian ones, though Germany publishes significantly more procedures on TED. Portugal demonstrated faster processes, albeit with a smaller sample size. With respect to harmonization, such comparisons can be performed via performance metrics, measuring both the similarity of legal frameworks and the duration of expected activities in the legal process flow.

The methodology has proven particularly effective in verifying compliance with newly imposed deadlines, such as those introduced by Law Decree 76/2020 and now incorporated into Legislative Decree 36/2023, which establish specific time-frames for different procurement procedures. To address procurement delays, several improvements can be targeted at critical phases. During the initial evaluation phase (80 days), digital pre-evaluation tools and simplified selection methods can streamline the process. For the negotiation period (22.5 days), digitization of procedures and reduced paper-based interactions are recommended. Contract duration efficiency (32.4-36 months) can be improved through early completion incentives, performance penalties, and project management software. These solutions align with legislative objectives while maintaining process quality.

6. Conclusions

The present work showed the feasibility, with the methodologies and initial results, of automated analysis in comparative law. The two examples presented in the paper, concerning both standardization and harmonization, have identified distinct challenges. An initial challenge lies in data quality, as the

accuracy and completeness of legal data are crucial for both methodologies. Furthermore, the integration of domain experts is indispensable, aligning with a human-in-the-loop approach. Despite technological advancements, legal expertise remains vital for providing context, judgment, and oversight. Finally, the interpretability of results presents a significant challenge. While AI and PM can identify patterns, understanding the underlying reasons for these patterns in the legal domain can be complex. As a potential future work, we aim to investigate further the adoption of additional features from openly available datasets to improve the process-oriented analysis.

Finally, it is essential to recognise the significance of the AI Act [37], particularly with regard to the transparency obligations imposed on AI systems. These requirements are fundamental to ensure that AI technologies operate in a manner that is accountable and understandable to all stakeholders but also play a key role in fostering public trust in AI innovations. By adhering to these transparency standards, organisations can demonstrate compliance with regulatory frameworks while promoting ethical and responsible AI development.

Acknowledgement

The research work in this article was partially conducted as part of the following projects: Circular Health European Digital Innovation Hub (CHEDIH) - Grant Agreement n. 101083745; FACILEX - Grant Agreement n. 10108963; GECOPAM and EXTRAI projects - Bando a Cascata PE FAIR Spoke 8.

References

[1] M. Siems, Comparative law, Cambridge University Press, 2022.

[2] G. Sileno, J. Spanakis, G. van Dijck (Eds.), Legal Knowledge and Information Systems - JURIX 2023: The Thirty-sixth Annual Conference, Maastricht, The Netherlands, 18-20 December 2023, volume 379 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2023. doi:10.3233/FAIA379.

[3] E. Sulis, L. D. Caro, R. Nanda, Introduction for computer law and security review: special issue "knowledge management for law", Comput. Law Secur. Rev. 52 (2024) 105949. doi:10.1016/J.CLSR.2024.105949.

[4] T. J. M. Bench-Capon, M. Araszkiewicz, K. D. Ashley, K. Atkinson, F. Bex, F. Borges, D. Bourcier, P. Bourgine, J. G. Conrad, E. Francesconi, T. F. Gordon, G. Governatori, J. L. Leidner, D. D. Lewis, R. P. Loui, L. T. McCarty, H. Prakken, F. Schilder, E. Schweighofer, P. Thompson, A. Tyrrell, B. Verheij, D. N. Walton, A. Z. Wyner, A history of AI and law in 50 papers: 25 years of the international conference on AI and law, Artif. Intell. Law 20 (2012) 215–319. doi:10.1007/S10506-012-9131-X.

[5] M. Dumas, M. L. Rosa, J. Mendling, H. A. Reijers, Fundamentals of Business Process Management, Second Edition, Springer, 2018. doi:10.1007/978-3-662-56509-4.

[6] E. Sulis, K. Taveter, Agent-Based Business Process Simulation - A Primer with Applications and Examples, Springer, 2022. doi:10.1007/978-3-030-98816-6.

[7] J. Munoz-Gama, N. Martin, C. Fernández-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, I. A. Amantea, R. Andrews, M. Arias, I. Beerepoot, E. Benevento, A. Burattin, D. Capurro, J. Carmona, M. Comuzzi, B. Dalmas, R. de la Fuente, C. D. Francescomarino, C. D. Ciccio, R. Gatta, C. Ghidini, F. Gonzalez-Lopez, G. Ibáñez-Sánchez, H. B. Klasky, A. P. Kurniati, X. Lu, F. Mannhardt, R. Mans, M. Marcos, R. M. de Carvalho, M. Pegoraro, S. K. Poon, L. Pufahl, H. A. Reijers, S. Remy, S. Rinderle-Ma, L. Sacchi, F. Seoane, M. Song, A. Stefanini, E. Sulis, A. H. M. ter Hofstede, P. J. Toussaint, V. Traver, Z. Valero-Ramon, I. van de Weerd, W. M. P. van der Aalst, R. J. B. Vanwersch, M. Weske, M. T. Wynn, F. Zerbato, Process mining for healthcare: Characteristics and challenges, J. Biomed. Informatics 127 (2022) 103994. doi:10.1016/J.JBI.2022.103994.

[8] R. Nai, E. Sulis, L. Genga, Enhancing e-learning effectiveness: A process mining approach for short-term tutorials, J Intell Inf Syst (2024). doi:https://doi.org/10.1007/s10844-024-00874-9.

[9] R. Nai, E. Sulis, E. Marengo, M. Vinai, S. Capecchi, Process mining on students' web learning traces: A case study with an ethnographic analysis, in: O. Viberg, I. Jivet, P. J. Muñoz-Merino, M. A. Perifanou, T. Papathoma (Eds.), Responsive and Sustainable Educational Futures - 18th European Conference on Technology Enhanced Learning, EC-TEL 2023, Aveiro, Portugal, September 4-8, 2023, Proceedings, volume 14200 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 599–604. doi:10.1007/978-3-031-42682-7_48.

[10] R. Whalen, Quantitative methods in comparative law, Elgar Encyclopedia of Comparative Law (2023).

[11] D. Audrito, E. Sulis, L. Humphreys, L. Di Caro, Analogical lightweight ontology of eu criminal procedural rights in judicial cooperation, Artificial Intelligence and Law 31 (2023) 629–652.

[12] E. Sulis, L. B. Humphreys, D. Audrito, L. Di Caro, Exploiting textual similarity techniques in harmonization of laws, in: S. Bandini, F. Gasparini, V. Mascardi, M. Palmonari, G. Vizzari (Eds.), AIxIA 2021 – Advances in Artificial Intelligence, Springer International Publishing, Cham, 2022, pp. 185–197.

[13] W. M. P. van der Aalst, Process Mining - Data Science in Action, Second Edition, Springer, 2016. doi:10.1007/978-3-662-49851-4.

[14] L. Potin, V. Labatut, P. Morand, C. Largeron, FOPPA: an open database of french public procurement award notices from 2010-2020, Scientific Data 10 - 303 (2023). doi:https://doi.org/10.1038/s41597-023-02213-z.

[15] A. J. Unger, J. F. dos Santos 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: J. Maranhão, A. Z. Wyner (Eds.), ICAIL '21: Eighteenth International Conference for Artificial Intelligence and Law, São Paulo Brazil, June 21 - 25, 2021, ACM, 2021, pp. 240–244. doi:10.1145/3462757.3466137.

[16] K. Rabuzin, N. Modrušan, S. Križanić, R. Kelemen, Process mining in public procurement in croatia, in: B. Lalic, D. Gracanin, N. Tasic, N. Simeunović (Eds.), Proceedings on 18th International Conference on Industrial Systems – IS'20, Springer International Publishing, Cham, 2022, pp. 473–480. doi:10.1007/978-3-030-97947-8_62.

[17] R. Nai, I. Fatima, G. Morina, E. Sulis, L. Genga, R. Meo, P. Pasteris, AI applied to the analysis of the contracts of the italian public administrations, in: 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, 2023, pp. 255–260. URL: https://ceur-ws.org/Vol-3486/100.pdf.

[18] R. Nai, E. Sulis, L. Genga, Automated analysis with event log enrichment of the european public procurement processes, in: T. P. Sales, J. Araújo, J. Borbinha, G. Guizzardi (Eds.), Advances in Conceptual Modeling - ER 2023 Workshops, Lisbon, Portugal, November 6-9, 2023, Proceedings, volume 14319 of *LNCS*, Springer, 2023, pp. 178–188. doi:10.1007/978-3-031-47112-4_17.

[19] W. M. P. van der Aalst, Process Mining - Discovery, Conformance and Enhancement of Business Processes, Springer, 2011. doi:10.1007/978-3-642-19345-3.

[20] M. J. Jans, M. Alles, M. A. Vasarhelyi, Process mining of event logs in auditing: Opportunities and challenges, SSRN 1578912 (2010). doi:dx.doi.org/10.2139/ssrn.1578912.

[21] R. Zaman, M. Hassani, On enabling GDPR compliance in business processes through data-driven solutions, SN Comput. Sci. 1 (2020) 210. doi:10.1007/S42979-020-00215-X.

[22] D. Gabbay, J. Horty, X. Parent, R. Van der Meyden, L. van der Torre, et al., Handbook of deontic logic and normative systems, College Publications, 2021, 2021.

[23] C. Sansone, G. Sperlí, Legal information retrieval systems: State-of-the-art and open issues, Inf. Syst. 106 (2022) 101967. doi:10.1016/J.IS.2021.101967.

[24] R. Mignone, I. Spada, C. Bonfanti, M. Colombino, G. Iacobellis, L. J. M. Zaharia, M. Molinari, I. A. Amantea, E. Sulis, L. D. Caro, G. Boella, Augmented reading and similar case matching: from legal domain experts' modus operandi to a computational pipeline, in: F. Toyoshima, M. Katsumi, G. Righetti, S. D. Giorgis, M. M. Hedblom, O. Kutz, B. Brodaric, M. Grüninger, T. Hahmann, D. D. Simon, M. Lange, H. Küçük-McGinty, A. S. Simon, R. Cameron, B. Gajderowicz,

D. Rosu, J. Hastings, D. Audrito, L. D. Caro, F. Grasso, R. Nai, E. Sulis, L. Bozzato, C. Shimizu, A. Zimmermann, R. Baratella, S. Borgo, S. de Cesare, T. P. Sales (Eds.), Proceedings of the Joint Ontology Workshops 2023 Episode IX: The Quebec Summer of Ontology co-located with the 13th International Conference on Formal Ontology in Information Systems (FOIS 2023), Sherbrooke, Québec, Canada, July 19-20, 2023, volume 3637 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3637/paper22.pdf.

[25] P. Casanovas, M. Palmirani, S. Peroni, T. Van Engers, F. Vitali, Semantic web for the legal domain: the next step, Semantic web 7 (2016) 213–227. doi:10.3233/SW-160224.

[26] R. Nai, E. Sulis, P. Pasteris, M. Giunta, R. Meo, Exploitation and merge of information sources for public procurement improvement, in: Machine Learning and Principles and Practice of Knowledge Discovery in Databases - International Workshops of ECML PKDD 2022, Grenoble, France, September 19-23, 2022, Proceedings, Part I, 2022, pp. 89–102. doi:10.1007/978-3-031-23618-1_6.

[27] R. Nai, E. Sulis, I. Fatima, R. Meo, Large language models and recommendation systems: A proof-of-concept study on public procurements, in: Natural Language Processing and Information Systems - 29th International Conference on Applications of Natural Language to Information Systems, NLDB 2024, Turin, Italy, June 25-27, 2024, Proceedings, Part II, 2024, pp. 280–290. doi:10.1007/978-3-031-70242-6_27.

[28] R. Nai, E. Sulis, R. Meo, Public procurement fraud detection and artificial intelligence techniques: a literature review, in: Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management, Bozen-Bolzano, Italy, September 26-29, 2022, 2022, pp. 45–56. URL: https://ceur-ws.org/Vol-3256/km4law4.pdf.

[29] N. Aletras, D. Tsarapatsanis, D. Preotiuc-Pietro, V. Lampos, Predicting judicial decisions of the european court of human rights: a natural language processing perspective, PeerJ Comput. Sci. 2 (2016) e93. doi:10.7717/PEERJ-CS.93.

[30] R. Nai, R. Meo, G. Morina, P. Pasteris, Public tenders, complaints, machine learning and recommender systems: a case study in public administration, Comput. Law Secur. Rev. 51 (2023) 105887. doi:10.1016/J.CLSR.2023.105887.

[31] R. Meo, R. Nai, E. Sulis, Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable AI... what's next?, in: S. Chiusano, T. Cerquitelli, R. Wrembel (Eds.), Advances in Databases and Information Systems - 26th European Conference, ADBIS 2022, Turin, Italy, September 5-8, 2022, Proceedings, volume 13389 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 25–34. doi:10.1007/978-3-031-15740-0_3.

[32] R. Nai, E. Sulis, R. Meo, ITH: an open database on Italian tenders 2016-2023, Scientific Data (2024) 1–15.

[33] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, I. Androutsopoulos, LEGAL-BERT: The muppets straight out of law school, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 2898–2904. doi:10.18653/v1/2020.findings-emnlp.261.

[34] F. De Caro, A. Del Monte, V. Scoppa, Corruption and public procurement: Evidence from italian municipalities, Journal of Public Economics 204 (2022) 104552. doi:10.1016/j.jpubeco.2021.104552.

[35] M. Popa, D. Voicu, X. Zhang, Uncovering corruption risks in public procurement, Government Information Quarterly 36 (2019) 206–219. doi:10.1016/j.giq.2018.12.001.

[36] D. Ovsyannikova, M. Fazekas, J. Wachs, Identification and analysis of legal frameworks in european public procurement, Regulatory Affairs Journal 15 (2020) 372–384. doi:10.1002/rag.1035.

[37] European Commission, Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, accessed: 2024-10-17.