Article

A New Research Agenda: How European Institutions Influence Law-Making in Justice and Home Affairs

Angela Tacea

FWO—Research Foundation Flanders, Institute for European Studies, Vrije Universiteit Brussel, Belgium;
E-Mail: angela.tacea@vub.be

Abstract
The article presents a dataset on the legislative procedure in European Justice and Home Affairs (JHA) and a new method of data processing. The dataset contains information on 529 procedures proposed between January 1998 and December 2017. For each of the legislative proposals, the dataset identifies the main elements of the legislative procedure (e.g., dates, types of procedure, directory codes and subcodes, actors, voting results, amendments, legal basis, etc.) and the changes introduced at each step of the legislative process from the text proposed by the European Commission to the final version published in the *Official Journal of the European Union*. This information has been gathered using text mining techniques. The dataset is relevant for a broad range of research questions regarding the EU decision-making process in JHA related to the balance of powers between European institutional actors and their capacity to influence the legislative outputs.

1. Introduction

How do EU institutional actors participate in and exercise influence on the law-making process? How did formal rule changes introduced by the successive EU treaties modify the capacity of institutional actors to determine legislative outputs? The academic literature answered these questions using mainly two methodological approaches. On the one hand, inspired by rational choice institutionalism, most studies used spatial models to understand how actors' preferences are transformed into decision outcomes. On the other hand, constructivist approaches to the EU decision-making process focused on single or comparative case studies. However, both methodological approaches bear important limitations for the understanding of the actual distribution of power in the EU decision-making process. This article presents a new research agenda to study the balance of power between EU institutional actors in the context of law-making procedures, by presenting a dataset on the legislative procedure in Justice and Home Affairs (JHA) and introducing a new text-mining method.

JHA has been often regarded as a specific EU policy field in the EU decision-making process due to its intergovernmental origins and to the longstanding disputes between Member States on institutional matters, which kept this policy field out of the traditional 'regulatory' mode of EU policymaking for a long time. However, starting with the Amsterdam Treaty, successive reforms of EU formal rules normalised the decision-making process in this field. Because of the evolution of the basic legal framework, but also because of the actors' conflicting perceptions in terms of substantive law, the area of JHA offers an ideal test case for assessing the role of institutional actors and their influence on the legislative outputs. The role of the European Parliament (EP) in JHA issues has increased significantly over time. When the third intergovernmental pillar was introduced by the

Maastricht Treaty, the Council of the European Union (hereafter referred simply as the Council) enjoyed a quasi-monopoly on decision-making and the EP had only a consultative role through its right to issue non-binding opinions. The EP was excluded from exerting any sort of influence on the legislative output (Crombez, 1996; Ripoll Servent, 2018a; Steunenberg, 1994). With the progressive communitarisation of JHA issues and the generalisation of the ordinary legislative procedure between 2005 and 2009, the role of the EP increased to the point that it now enjoys equal legislative rights with the Council. Consultation of the EP still applies for the adoption of measures on administrative cooperation in the fields of policing and criminal law and unanimity has been retained for issues relating to passports, family law, the European public prosecutor (with the EP having a power of consent) and operational police cooperation. Despite those specificities, the generalisation of the ordinary legislative procedure could reasonably be interpreted as the EP exerting an influence equal to that of the Council on the legislative outputs.

However, the academic literature offers contrasting findings when it comes to assessing the impact of rule change on the capacity of institutions to determine the legislative outputs. On the one hand, spatial models suffer from a misinterpretation and misrepresentation of the EU legislative procedures (Crombez & Vangerven, 2014). Due to the equivocal nature of formal rules, the relative power of the Council, the European Commission and the EP has been an element of intense debate, even among scholars adopting very similar theoretical and methodological approaches (Thomson & Hosli, 2006). For example, while most formal studies argued that the power of the EP increased with the generalisation of the co-decision procedure (Mcelroy, 2006; Steunenberg, 2000; Thomson, 2011), some scholars claimed that the legislature has been weakened by this constitutional change because the EP has lost its ability to act as a conditional agenda setter (Tsebelis & Garrett, 2000). In addition, formal models are based on a close reading of the EU treaties to precisely specify the hypothesised policy process or form of the game. They view institutional environments as static and institutional preferences as stable. Or research has shown that institutional arrangements are inherently dynamic, and actors might not behave according to the formal rules (Kleine, 2013). They engage in informal practices to avoid deadlock situations (Farrell & Héritier, 2003). On the other hand, case studies on the JHA proposals suggest that the formal empowerment of the EP did not materialise in practice, because formal rule changes did not result in substantive policy change (Trauner & Ripoll Servent, 2015). After the generalisation of the co-decision procedure, the EP, which has been known for its extreme positions, tended to be more moderate and to favour positions at the centre of the political spectrum (Ripoll Servent, 2018). Contrary to what has been suggested by Tsebelis and Garrett (2000), the influence of the EP on legislative outputs is limited and

the Council still dominates the legislative procedure and policy positions made jointly by member states generally matter more than the policy positions of the EP (Laloux & Delreux, 2018; Thomson, 2011).

To overcome these different and sometimes contradictory findings, there is a need to connect case studies focusing only on a few salient JHA proposals to the wider literature on the role and influence of EU institutional actors in each type of legislative procedure, overcoming the fragmentation and overspecialisation of studies of the EU individual policy areas. The dataset presented in this article contains information on the degree of change JHA proposals undergo during the legislative procedure. It is innovative as it examines the role of the main institutional actors in determining the legislative output, while capturing the specificities of a particular EU policy field. First, the dataset maps the whole law-making process in JHA and the actors involved. Second, it identifies the changes JHA proposals undergo at each step of the legislative process from the text proposed by the European Commission to the final version published in the *Official Journal of the European Union*. Since the dataset includes information on the legislative process in JHA from 1998 to 2017, it can be placed in the broader context of the impact of formal rule change on substantive democratic governance in the EU. Indeed, the time frame starts with the Treaty of Amsterdam, which marked an important steppingstone for the EP by shifting some of the third pillar measures (immigration, asylum, border controls, visa and civil law cooperation, with the exception of family law) to the first pillar and subjecting them to co-decision; the time span ends in 2017, seven years after the entry into force of the Lisbon Treaty, which confirmed the EP as a full co-legislator in JHA. Subsequently, the article introduces an innovative method to study the EU legislative procedure. Though this study is not the first to apply text mining techniques to the EU decision-making process (see, among others, Casas et al., 2020; Cross & Hermansson, 2017; Gava et al., 2020), its contribution to this developing literature is two-fold. First, it is the first study to apply this method to all actors involved in the law-making process and at all stages of it. Second, unlike other studies, it follows the open data movement and data is made publicly available. Such a method combines flexibility with accuracy and replicability and offers a more fine-grained measure of legislative change than the number of amendments or the number of words changed. The next section describes the dataset. The article then presents a new research agenda on the balance of power in JHA by discussing several research questions that can be answered using the dataset and illustrate them with some examples. The final section addresses the limitations of the dataset.

2. The Dataset

This section offers a description of the dataset developed in the framework of the AFSJ-Pol-Lex-Track research

project and of the methodology used to assemble it. The dataset contains two types of information: (1) quantitative information related to each legislative procedure and (2) qualitative text data about each step of the legislative procedure.

The first component of the dataset contains general information about the legal act (legal basis, title of the legislative act, inter-institutional code, CELEX number, type of procedure, type of act, directory codes and subcodes, total duration of the procedure in number of days, etc.), as well as information about the procedure in each institution (EP votes, names and party affiliation of the rapporteurs and of the rapporteurs for opinion, EP position at 1st reading, 2nd reading, 3rd reading, dates of the Council's political agreement, Council position at 1st reading, 2nd reading, 3rd reading, number of points A and B on the agenda of the Council, position of the European Commission on EP amendments at each reading, etc.). The main source to extract the data is the European Commission's website (EUR-Lex), which contains all documents printed in the *Official Journal of the European Union*. Though EUR-Lex provides the stages of the legislative procedure, an accurate picture can be obtained only by corroborating all the available sources. Thus, data extracted from EUR-Lex is complemented by data extracted from the EP Legislative Observatory (OEIL) and the Council's Document Register, using the inter-institutional code (e.g., 2016/0412/COD) as the common reference number for all European institutions. Between 1998 and 2017, EU institutions adopted 746 legal acts and 101 international agreements in JHA. From those 847 normative acts, I removed all non-binding legal acts (resolutions, opinions) and other instruments (EU institutions' internal regulations, EU action programmes, etc.) and codification procedures, which are processes of bringing together a legal act (or several related acts) and

all its amendments into a single new act. I was thus left with N = 536 procedures. Table 1 in the Supplementary File 1 offers an overview of the information the dataset provides about the legislative procedure in JHA (the dataset and the detailed codebook are available on a GitLab repository: https://gitlab.com/shoricitza/afsj-pol-lex-track-quantitative-dataset). Table 1 offers some descriptive statistics about the first component of the dataset.

The second component of the dataset contains text data about each step of the legislative procedure. Text data (PDF/HTML/XML) is also extracted from EUR-Lex, OEIL, and the Council's Document Register. PDF/HTML/XML files extracted were converted to plain text and pre-processed to make them comparable. The text is structured following the standard legislative structure (see Supplementary File 2). Several documents, mostly those related to the informal trilogue negotiations, are not publicly available. Individual requests for documents have been submitted to the EP and the Council. To understand which stages and text of the legislative procedures to include in the analysis, I conducted six exploratory interviews with senior officials from the EP directorate for legislative acts, the EP unit for reception and referral of official documents, the Council's legal service—quality of legislation—legislative acts/planning, the legal data processing group of the Council, the Council's information services and the Directorate-General (DG) for European Parliamentary Research Services. Following these interviews, I included in the dataset four main types of legislative steps: the European Commission (amended) proposal, the EP committee and plenary reports, the Council's negotiation mandate and/or common position, and the final act signed by the Presidents of the EP and of the Council. Legal linguists sometimes make substantial changes to

Table 1. Descriptive statistics for the legislative procedure in JHA.

Type of procedure	Status of the procedure			Number of readings			Average procedure duration (days)	Number of points A/B on the agenda of the Council	
	Adopted	Withdrawn	Pending	1	2	3		A	B
Assent (APP)	4	1					853,75	2	0
Assent (AVC)	3						211,33	3	0
Consultation	199	23	1				439,26	202	181
Co-decision	55	6		42	18	1	773,29	63	69
Non legislative procedure	74		20				474,64	16	2
Ordinary legislative procedure	85	3	22	99	11		687,03	22	56
Special legislative procedure	5	2					382,8	1	4

the act signed by the Presidents of the EP and the Council, which biases the analysis of the modifications institutional actors introduce to the legislative proposal and the political negotiations between the three institutional actors involved in the legislative process. For this reason, the final act published in the Official Journal has been excluded from the analysis. For those procedures where trilogues took place, I included an additional step represented by the four-column documents of the trilogues and the COREPER letter confirming agreement.

3. Measuring the Degree of Change the JHA Proposals Undergo during the Legislative Procedure Using Text Mining

Past studies have relied on different types of data and various methodologies to assess the balance of power between EU institutional actors. Datasets have been established using the evaluation of experienced practitioners of EU policies in these institutions (Neuhold & Dobbels, 2015; Thomson, 2011, 2015) or the quantification of the EP amendments (Kreppel, 2002; Tsebelis et al., 2001). Different methodological approaches have also been used to analyse these data, ranging from formal modelling (Costello & Thomson, 2013; Selck, 2006) to inferential statistics (König, 2008; Kreppel, 2002). While these studies shed light on the distribution of power among the EU institutions, they all bear important empirical limitations: (1) practitioners may not have the same understanding of power as academics, (2) parliamentary amendments do not reflect the informal negotiations among actors and do not distinguish between formal and substantive changes, and (3) formal models and trilogue studies do not capture all the stages of the legislation-making process. In addition, they tend to focus only on one institution, with most of the studies analysing the EP (Kreppel, 2002) or the Council (Thomson, 2011, 2015), or only on certain stages of the legislative procedure, either the formal or the informal negotiations (for some exceptions see Laloux & Delreux, 2020; Thomson, 2015).

To overcome the empirical limitation of past studies, I use a new machine-learning based approach to analyse texts. Text mining techniques have the advantage of capturing both quantitative information, such as the length of laws in words or the number of amendments/modifications. But also, the substantive content of legislation can be analysed, which would otherwise require extensive human input that is not viable for large quantities of legislative text. At the same time, compared to other methodologies, text mining techniques can be easily replicated and applied in a variety of contexts. Recent studies have used text analysis methods to evaluate the impact of formal institutional settings on amendment capabilities. For example, Cross and Hermansson (2017) use minimum edit distances to show that there are significantly more successful amendments to a European Commission proposal under co-decision compared to the consultation pro-

cedure. Gava et al. (2020) use a dissimilarity index to assess the capacity of the Swiss parliament to amend bills. Peterson employs vector word embeddings to analyse Congressional modification of legislation (Peterson, 2017). Laloux and Delreux (2020) compute the percentage of words that appear at each phase of the legislative process to trace the origin of EU legislation.

In line with these studies, I rely on text reuse methods to assess the degree of change the JHA proposals undergo at each legislative step and the extent to which the changes proposed by actors are included in the final adopted text. The approach of text reuse methods is based on the idea that similarity between texts can be assessed by looking at how much text is common to two versions of the proposal. Accordingly, I compared the full texts adopted at the different stages of the legislative proposal with the final adopted version. The comparisons are done in pairs, two at a time. More precisely, I compare the proposal of the European Commission with the final adopted text; I then compare the report of the EP with the final adopted text etc. I use the FuzzyWuzzy Package in Python to assess the extent to which institutional actors modify the text. I calculate a similarity index between an actor's position at different times in the legislative process and the final legislative output. A detailed presentation of the FuzzyWuzzy Package is provided in the Supplementary File 1. The similarity index varies between 0 percent and 100 percent, where 0 means that the text adopted by a specific actor is totally different from the final adopted text, and 100 indicates that the text is exactly the same as the final adopted text. The index can be interpreted as the rate of change between an actor's position and the final legislative output and is calculated for each component of the legislative proposal (preambles, articles, annexes).

To visualise the evolution of JHA legislative procedures, I developed a web application (https://shoricitza.gitlab.io/afsjlexpol). For each legislative procedure, the web application provides a visualisation of the text adopted at each stage of the legislative procedure (e.g., the text proposed by the European Commission, the positions of the EP at each reading, trilogues four column documents, the positions adopted by the Council at each reading, etc.) and the similarity index between the text adopted at each step of the legislative procedure and the final adopted one. All the similarity indexes can be freely downloaded in .csv format from the web application for each legislative proposal.

4. Research Questions That Can Be Answered Using the Dataset

In this section, I discuss some of the research questions to which the dataset is relevant, by providing some examples, and identify some questions that could be further developed. The dataset can be used to address the discrepancy between formal and substantive democratic governance by examining the link between actors' formal

power and their influence on legislative outputs. It does so by offering a broad understanding of the legislative procedure, without losing the specificities of the JHA policy area.

4.1. Do Actors Make Use of Their Formal Prerogatives or Not?

At the aggregate level, similar to both formal models and JHA case studies literature, the dataset suggests a limited role of the EP in the consultation procedure. On the one hand, between 1998 and 2017, the EP rejected 15 percent of the proposals initiated under consultation—most of them being member states' initiatives. The Council completely ignored the opinion of the Parliament and adopted the texts. Nonetheless, contrary to the conclusions of the literature on JHA, the EP position in consultation is not radically different to that of the Council. As shown in Table 2, the average similarity index between the final adopted text and the text adopted by the EP is 88.5 percent.

   A few exceptions should be mentioned, such as the Council Directive relating to the conditions in which third-country nationals shall have the freedom to travel in the territory of the member states for periods not exceeding three months (2001/0155/CNS) or the Blue Card directive. However, these examples tend rather to be exceptions. The dataset offers the possibility to go beyond a general overview of the legislative proposal and understand the influence of actors, article by article, thus offering a more nuanced picture of actors' behaviour. The legislative proposal on giving temporary protection in the event of a mass influx of displaced persons offers an interesting example. On a general level, the positions of the three actors are rather similar with a similarity index of 86.63 percent between the proposal of the European Commission and the final text adopted by the Council and of 85.82 percent between the text adopted by Parliament and the final text adopted by the Council. However, at the individual level of the articles, the picture is more nuanced. On the one hand, the EP introduced several substantive amendments that were completely ignored by the Council. For example, it introduced a new paragraph in article 8 granting persons enjoying temporary protection access to their territory and amended article 13 to better protect the right to family reunification, neither of which was included in the final text. As can be seen, the similarity index between the

position of the EP, both at the committee and at the plenary level, and the final text adopted by the Council, is only 53 percent, showing a limited influence of the EP on the final text:

   Article 13: Similarity index and article size

   sim_ Commission's proposal: 86%

   article_size_ Commission's proposal: 2833

   sim_ EP committee report: 53%

   article_size_ EP committee report: 1030

   sim_ EP position at 1st reading: 53%

   article_size_ EPs position at 1st reading: 1032

   article_size_ Adoption by Council: 1045

On the other hand, there are instances where the Council partially retained the amendments proposed by the EP. For example, the EP substantially modified article 18, which affirmed the incompatibility of temporary protection with the status of asylum, to offer more guarantees to asylum seekers. Here, however, the Council partially included the modifications suggested by the EP in the final adopted text. The similarity index between the EP position and the final adopted text is 86 percent:

   Article 18: Similarity index and article size

   sim_ Commission's proposal: 59%

   article_size_ Commission's proposal: 472

   sim_ EP committee report: 86%

   article_size_ EP committee report: 662

   sim_ EP's position at 1st reading: 86%

   article_ size_EP's position at 1st reading: 658

   article_size_ Adoption by Council: 363

Table 2. Descriptive statistics consultation procedure (%).

	COM proposal	EP committee report	EP plenary	Adoption by the Council
Min	57	57	57	100
Max	100	100	100	100
Average	87.76	87.35	88.58	100
St dev	6.47	6.85	7.04	110
N	248	248	248	248

4.2. Did the Generalisation of the Co-Decision Procedure Result in an Increased Influence of the EP?

When it comes to the co-decision procedure, the dataset suggests that the distribution of power between the three institutional actors is more balanced compared to the consultation procedure. In this sense, the dataset tends to support the conclusions of the empirical studies on JHA, which argue that the EP favoured compromise with the Council, even when this went against its own preferences, and only occasionally used its new prerogatives to impact on the development of JHA policies (Lopatin, 2011; Ripoll Servent, 2013; Trauner & Ripoll Servent, 2015).

Table 3 gives an overview of the co-decision procedure before the entry into force of the Lisbon Treaty. Table 4 does the same for the time period after the Lisbon Treaty.

Two preliminary findings can be drawn from the tables below. First, after the Lisbon Treaty, both the EP and the European Commission proposed texts that were not extremely different from the final adopted text, showing thus a more pragmatic negotiation strategy. Second, as pointed out by the JHA literature, the EP tended to compromise more with the Council after the entry into force of the Lisbon Treaty, compared to the period before. However, this does not mean that the influence of the EP is limited, substantive parliamentary amendments being incorporated in the final text. The legislative proposal on combating fraud and counterfeiting of non-cash means of payment (2017/0226/COD), which was randomly selected from the dataset, offers a case in point. The proposal aimed to update the Council Framework Decision 2001/413/JHA on combating fraud and counterfeiting of non-cash means of payment in order to adapt it to the new challenges and technolog-

ical developments such as virtual currencies and mobile payments. In this sense, it sought to establish a framework to deal effectively with non-cash payment fraud. At the aggregate level of the proposal, all three institutions entered the legislative procedure with positions that were rather different compared to the final adopted act. Though important differences can be noticed on each article, an agreement on the text was reached only during the third and last trilogue negotiation.

4.3. How Do Formal Rules and Informal Practices Affect the Distribution of Power?

The data shows the importance of informal negotiations. Indeed, neither the Council, nor the EP, nor the European Commission had a clear determinant role in the final output in the above-mentioned example. For example, while the final adopted article 19 is similar to the position of the Council in the third trilogue, article 20 reflects rather the position of the EP. The aggregate similarity indexes confirm the same trend, and the average similarity index of the EP position passes at 99.71 percent after trilogue negotiations. By providing information on all stages of the legislative procedures, the dataset integrates informal practices into the study of EU legislative process. Existing research stressed that the informalisation of the legislative procedure has become particularly prominent in co-decision/ordinary legislative procedure (Brandsma, 2019; Reh et al., 2013; Roederer-Rynning & Greenwood, 2015). Trilogues have become the main mechanism for inter-institutional legislative negotiations, and they can persistently and systematically depart from formal rules (Brandsma, 2019; Farrell & Héritier, 2003; Kleine, 2013; Reh et al., 2013; Thomson, 2015). However, despite an increased academic interest in informal practices since their emergence in the early

Table 3. Descriptive statistics co-decision procedure before Lisbon (%).

	COM proposal	EP committee report 1st reading	EP plenary 1st reading	Council position	COM amended proposal	EP committee report 2nd reading	EP plenary 2nd reading	SIGN EP-Council
Min	35	30.22	46.53	57.73	45.58	57.73	51.52	100
Max	90.84	94.15	97.65	99.47	92.15	99.61	99.53	100
Average	84.60	85.41	86.93	77.54	74.89	82.26	86.47	100
St dev	13.44	13.11	12	20.83	21.29	19.95	19.05	0
N	61	61	61	19	19	19	19	61

Table 4. Descriptive statistics co-decision procedure after Lisbon (%).

	COM proposal	EP committee report	EP plenary	General approach	SIGN EP-Council
Min	75.48	75.29	99.7	87.70	100
Max	93.93	100	100	96.86	100
Average	92.74	95.81	99.71	94.05	100
St dev	5.08	6.54	1.16	6.3	0
N	110	110	110	110	110

2000s, with very few exceptions (Laloux & Delreux, 2020; Thomson, 2015) research into trilogues focused on each institution, with most of the studies analysing the EP, failing to integrate informal practices into the whole EU legislative decision-making process. This is mainly due to the lack of data. Using the dataset, future research might examine to what extent and in which direction actors modify their position during the informal negotiations.

The data also shows a surprising phenomenon. Though the Lisbon Treaty increased the formal powers of the EP by making it a full co-legislator in JHA issues, its role is limited by the spectacular increase in non-legislative procedures (NLP) which can be noticed in this area after the entry into force of the Lisbon Treaty. From 2010 until 2017, 86 NLPs, representing 40 percent of the adopted JHA acts, have been initiated compared to only 10 before 2010. Most of those NLPs concern the negotiation and conclusion of formal agreements with third countries, measures in the domain of family law, measures in the field of criminal procedural law not already foreseen by the Treaty, as well as EU/Schengen common policy on visas. The Lisbon Treaty massively strengthened the role of the EP in the external dimension of the AFSJ, allowing it to ratify international agreements in internal security with co-decision or consent being required for almost all acts. However, in practice the NLPs initiated between 2010 and 2017 required only consultation with the EP. Those procedures were adopted either on the basis of the Treaty on the Functioning of the European Union (TFEU; Art *81*(3)), such as proposals authorising different EU member states to accept the accession of third countries to the 1980 Hague Convention on the Civil Aspects of International Child Abduction, or on Art *78*(3) TFEU, which provides a specific legal basis to deal with emergency situations at the external borders. This high increase of NLPs proves that formal rules provide critical openings for agency and the European Commission choses strategically an institutional rule that limits the scope of the legislative powers the EP gained with the extension of the co-decision to the JHA.

4.4. Which Factors Explain the Influence of Institutional Actors on Legislative Outputs?

There is nonetheless scope for further inquiry into the causes of actors' success in influencing the final legislative output. For example, the dataset can be used to understand how much the EP has a voice and why. The dataset suggests that formal institutional change did not have much impact on the capacity of the EP to influence legislative output. Other variables might be at play. As such, indicators measuring both actors' formal resources (e.g., types of procedure, member states' voting weights in the Council, voting rules in the Council and the Parliament, legal nature of the acts, etc.) as well as their informal weight in the decision-making process (e.g., technical expertise of the rapporteur, congruence

between the rapporteur and presidency of the Council, policy expertise of the DGs, etc.) could be used. At the same time, the potential to influence legislative outputs might be related to the incentives actors face to mobilise their power. Those incentives are determined by policy attributes (e.g., degree of Europeanisation, technical complexity of the proposal, salience of the policy for public opinion, degree of conflict/consensus, level of unanimity in the EP and/or the Council), as well as by the relations between actors in the context of the legislative proposal (share of policy core beliefs).

4.5. Going beyond the Traditional Methodological Approaches

Lastly, the dataset offers a test case for the relevance of text reuse methods to study the EU law-making and decision-making processes more broadly. Identifying substantive differences in the legislative proposals, beyond that of simply counting the number of words or amendments, gives us more insights into the nature of modifications introduced by each institutional actor. Consider for example the activity of the EP's committee compared to that of the plenary. The basic logic is that if the final version of the adopted text is similar to the plenary version and different from the committee one, the plenary is influential. In other words, the fact that an actor makes significant changes to the legislative proposal may be considered at first sight as evidence of that actor's influence on the final legislative output. However, the fact that the plenary version is more similar to the final adopted act than the committee version could mean that the EP modified the proposal to ensure that the proposal is adopted at first reading. In co-decision, most of the time, the text the EP adopts in the plenary is the text that results from the informal trilogue negotiations. By analysing the evolution of the legislative proposal during the legislative process, we can clearly indicate which actor is responsible for the bulk of textual modifications of the legislative act and at which stage (formal or informal). Contrary to formal modelling which interprets and models the power of institutional actors mainly based on their formal treaty prerogatives, text mining offers an accurate empirical measure of the influence of actors on legislative outputs. At the same time, contrary to small-N case studies, which generalise their conclusions from a very limited number of cases, text mining techniques allow the study of very large numbers of legislative texts. Moreover, by comparing one version of the text to another, such methods provide insights into whether the changes made by one organ of an institution reverse the modifications introduced by another organ of the same institution. For example, if the plenary of the EP reverses the changes introduced by the committee, or if the COREPER reverses the changes introduced by the Council's working groups. Thus, applying text mining techniques to EU law-making provides an understanding not only on the balance of power

between the institutions, but also on the distribution of power between different organs of the same institution. There is however scope for further inquiry into matching the content of actors' modifications and the content of the final adopted act. For example, research on different ASFJ policy areas has shown that contrary to scholarly expectations, despite the empowerment of the EP, the rationale of providing 'security'—in all its expressions—remains dominant (Trauner & Lavenex, 2015, p. 220). Or as previously mentioned this research is based on case studies of salient EU proposals, thus it is still unknown if this conclusion holds true for more 'routinised' legislation. One way of filling this gap is to use 'active learning' which is a supervised learning approach, to match actors' substantive modifications with the final adopted act (Casas et al., 2020). The logic behind 'active learning' is to identify a small number of cases where the positions of the Parliament, the Council and the European Commission match the final adopted act and identify the policy core beliefs (e.g., protection of human rights vs security, data protection vs data processing, border control vs integration, etc.) and then assign these different dimensions to each paragraph of the legislative text. A trained classifier can be used to predict the share of actors' substantive modifications in the adopted legislative proposal for the whole text corpus.

Thus, the dataset and the use of text mining techniques provide a valuable source for scholars interested in EU legislative procedures in general, as well as intra-institutional negotiations in different policy fields and the formal and de facto legislative influence of the European Commission and the EP in particular. Nonetheless, as with any kind of dataset and method, it also has limitations.

5. Limitations of the Dataset

In this section I address two main criticisms that can be directed to the dataset, which are linked to its limitations. I also draw attention to the problems scholars might face when working with EU text data.

The first criticism is that the dataset treats EU institutions as homogenous actors, leaving no room for tracing the political games inside each institution and the internal political dynamics, all of which might have important consequences for inter-institutional relations. Related to this, another criticism is that official documents measure only the revealed/strategic preferences of actors. In response, I first point out that the qualitative dataset offers different proxies for the identification of politicisation inside the institutions. For the EP, the dataset contains information on the political affiliation of rapporteurs and shadow rapporteurs, the degree of unanimity in the EP (votes for, against, and abstentions), the number of parliamentary amendments and the position of the European Commission on each parliamentary amendment. For the Council of Ministers, the dataset provides the number of items A (proposals for which an agreement has been reached at the COREPER level) and B (proposals for which no agreement has been reached or politically sensitive issues) on the agenda of the Council, which reflect the degree of consensus/conflict a particular proposal raises (Novak et al., 2020). The dataset also provides information on the nature of the legislative proposal. For example, the controversial/uncontroversial nature of a legislative proposal can be deduced from the timespan between the European Commission's proposal and the EP committee report/Council's position, or the number of readings. The technical complexity of a policy can be measured by the number of DGs involved (see Laloux, 2021; Senninger et al., 2020).

However, the text data does not allow identification of member states' positions within the Council, nor the origin of parliamentary compromise amendments. The reason why I disaggregated this text data is the lack of systematisation of information provided by the EU institutions. If the Council's outcomes of proceedings provide the position of member states on certain issues in the footnotes, it does not do so in a systematic manner. Sometimes these positions are completely lacking, other times they are concealed due to the sensitive nature of the issues discussed. Moreover, as shown by Brandsma et al. (2021, p. 19), the shift towards bilateral forms of mandating and the change in Council's practice in 2014 resulted in no mention of the member states' positions in the footnotes. Thus, it becomes impossible to trace this information using an automated text data extraction. Tracing the origin of parliamentary compromise amendments is even more problematic than that of member states' positions in the Council. All senior officials of the EP who were interviewed pointed out the private nature of the political negotiations and the difficulty of assessing how much of the amendment has been incorporated into the final compromise amendment. Though some researchers (Ripoll Servent & Panning, 2021) have tried to estimate the level of incorporation of EP amendments into the final agreement, their measure is rather crude. Obviously, this limitation makes the dataset of little relevance for researchers interested in understanding the ideological battles inside the EU institutions and the left-right, pro/anti-integration or the GAL-TAN cleavages. More qualitative methods, susch as interviews with the relevant actors, can be used to capture those dynamics.

Another criticism is that in co-decision/ordinary legislative procedure, the positions of the Council and of the EP before trilogue negotiations refer to different types of documents and, therefore, cannot be compared, at least not in the same way. For example, while for some legislative proposals, the general approach is retained as the Council's position before negotiations, for others the compromise text of the Presidency or the political agreement are used. While the compromise text reflects agreement in the Council Working Parties, the general approach is the preliminary agreement on the text, as agreed at ministerial level. The same can be said about

the EP, where different documents such as committee draft reports and negotiation mandates, are used.

The first response to this criticism is that the different terminologies used do not necessary reflect different types of documents. Before January 2012 and the publication of the Council note on the *Terminology to be used in Council and COREPER agendas for legislative items under the Ordinary Legislative Procedure* (5084/12), the terms were not used consistently by the different organs of the Council. As acknowledged by one of the senior officials in the Council's legal service—quality of legislation—legislative acts/planning, this not only created confusion, but also had unplanned procedural consequences as the wrong files were sometimes used during the interinstitutional negotiations. To overcome this difficulty, all files were manually checked to verify that they were reflecting the agreement of the Council on the legislative proposal pending the EP vote. After 2012, the Council made an effort to clarify and systematise the terminology used. In the EP, though the annex XXI of the *Code of Conduct for Negotiating in the Context of the Ordinary Legislative Procedures of the EP's Rules of Procedure* adopted in 2009, established some general principles regarding the preparation of trilogue negotiations, there was no clear procedure when it came to the document used by the negotiation team. The mandate could have been the committee legislative report, or the amendments adopted in plenary for first-reading negotiations. Eventually, this system was reformed in 2012 and the EP mandate is based either on a report adopted in committee or the position adopted in Plenary and clearly identified as negotiation mandate (as in Art. 71 of the *Rules of Procedure of the European Parliament*). As in the case of the Council's position, all parliamentary documents have been manually checked.

The second response to this criticism regarding comparability of the documents is that the positions are comparable in that they are used by the institutions themselves as negotiation mandates. Independently of the typology or of the level at which the documents were adopted (committee vs plenary; working groups vs ministers), I used the documents as designated by each institution as its position before entering the interinstitutional negotiations. Using this information, it is possible to identify patterns of legislative change. For example, in consultation, the vast majority of parliamentary positions are very similar to that of the European Commission (average similarity index = 95 percent). These results might suggest that the European Commission and the Parliament act together as integration-minded actors in JHA, contrary to member states, which favour a more intergovernmental approach to JHA.

Lastly, scholars who use automated data collection and text mining methods to analyse EU text data must be aware of the poor quality of the data provided by the EU institutions. First, there are discrepancies between the data repositories/web services of each institution. Though each institution provides the legislative proce-

dure stages, an accurate picture can be obtained only by corroborating all the available sources. Indeed, data contained on EUR-Lex is reliable and complete only for what concerns the European Commission. The same is true for the OEIL and for the Council's document register. In addition, data provided by the Council is unstructured, which means that different methods have to be used to extract the data. I used web scrapping and corroborated the results with the open data provided by the Council. Second, several documents for the EP and the Council, for the time period 1999–2003 (approx. 150 procedures), correspond in practice to other procedures than those indicated. For example, for the proposal 2000/0304(CNS) *Fight against Organised Crime: Financial Support, Programme for the Prevention (Hippocrates)*, the link of the EP committee on OEIL refers to a completely different proposal. Thus, data should be manually validated to ensure its validity. Third, text data is not similarly structured for each institution. For example, while the European Commission provides the full (amended) text, the EP only provides a list of amendments for the position adopted at the committee level and, sometimes, though not systematically, the consolidated text for the position adopted by the plenary. The Council provides the whole text modified using bold, strikethrough or underlining. Those discrepancies require important pre-processing efforts because to be comparable the text should have the same structure.

6. Conclusion

Despite these different limitations of the dataset and challenges raised by automatic text analysis of EU data, the AFSJ-Pol-Lex-Track dataset offers a valuable source of information for assessing the interplay between actors' formal power and policy outcomes. In this sense, it becomes possible to identify and study the law-making patterns, dynamics and issues within the JHA law-making procedure, across time and across policy sub-fields. It does so by combining in a single enhanced environment, the ability to gather, analyse and relate different stages of the law-making process, the different texts proposed and the negotiations between the European Commission, the EP and the Council. By individualising the text versions adopted at each legislative stage, the visualisation application and the similarity indexes enable the identification of each institutional actor involved in the legislative process and of the legislative modifications it introduces to the text: Which part of the final text originates from the European Commission? Which part originates from the EP and which from the Council? Who is at the origin of the modifications that change the draft legislative proposal substantially?

The similarity indexes offer a more quantitative longitudinal measure of the capacity of actors to determine policy outputs. There is scope for further inquiry into the nature of modifications introduced during the legislative process. For this reason, active learning (supervised or

unsupervised) techniques can be used to match actors' substantive modifications with the final adopted act and to identify the policy ideas (e.g., protection of human rights vs security, data protection vs data processing) that are incorporated into the final adopted text. Beyond the JHA policy area, the methods presented in this article to collect data on the EU legislative procedure and to analyse it can be applied to any EU policy field, thus providing a new perspective in EU legal and political science studies.

Conflict of Interests

The author declares no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Brandsma, G. J. (2019). Transparency of EU informal trilogues through public feedback in the European Parliament: Promise unfulfilled. *Journal of European Public Policy*, *26*(10), 1464–1483.

Brandsma, G. J., Dionigi, M. K., Greenwood, J., & Roederer-Rynning, C. (2021). Trilogues in Council: Disrupting the diplomatic culture? *Journal of European Public Policy*, *28*(1), 10–31.

Casas, A., Denny, M. J., & Wilkerson, J. (2020). More effective than I thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process. *American Journal of Political Science*, *64*(1), 5–18.

Costello, R., & Thomson, R. (2013). The distribution of power among EU institutions: Who wins under codecision and why? *Journal of European Public Policy*, *20*(7), 1025–1039.

Crombez, C., & Vangerven, P. (2014). Procedural models of European Union politics: Contributions and suggestions for improvement. *European Union Politics*, *15*(2), 289–308.

Crombez, C. (1996). Legislative procedures in the European Community. *British Journal of Political Science*,
26(2), 199–228.

Cross, J. P., & Hermansson, H. (2017). Legislative amendments and informal politics in the European Union: A text reuse approach. *European Union Politics*, *18*(4), 581–602.

Farrell, H., & Héritier, A. (2003). Formal and informal institutions under codecision: Continuous constitution-building in Europe. *Governance*, *16*(4), 577–600.

Gava, R., Jacquet, J. M., & Sciarini, P. (2020). Legislating or rubber-stamping? Assessing parliament's influence on law-making with text reuse. *European Journal of Political Research*, *60*(1), 175–198.

Kleine, M. (2013). *Informal governance in the European Union. How governments make international organizations work*. Cornell University Press.

König, T. (2008). Why do member states empower the European Parliament. *Journal of European Public Policy*, *15*(2), 167–188.

Kreppel, A. (2002). Moving beyond procedure. An empirical analysis of European Parliament legislative influence. *Comparative Political Studies*, *35*(7), 784–813.

Laloux, T. (2021). Why defend something I don't agree with? Conflicts within the Commission and legislative amendments in trilogues. *Politics and Governance*, *9*(3), 40–51.

Laloux, T., & Delreux, T. (2018). How much do agents in trilogues deviate from their principals' instructions? Introducing a deviation index. *Journal of European Public Policy*, *25*(7), 1049–1067.

Laloux, T., & Delreux, T. (2020). The origins of EU legislation: Agenda-setting, intra-institutional decision-making or interinstitutional negotiations? *West European Politic*s. Advance online publication. https://doi.org/10.1080/01402382.2020.1836861

Lopatin, E. (2011). The changing position of the European Parliament on irregular migration and asylum under co-decision. *Journal of Common Market Studies*, *51*(4), 740–755.

Neuhold, C., & Dobbels, M. (2015). Paper keepers or policy shapers? The conditions under which EP officials impact on the EU policy process. *Comparative European Politics*, *13*(5), 577–595.

Novak, S., Rozenberg, O., & Bendjaballah, S. (2020). Enduring consensus: Why the EU legislative process stays the same. *Journal of European Integration*. Advance online publication. https://doi.org/10.1080/07036337.2020.1800679

Peterson, A. J. (2017). *Understanding politics through the machine learning of texts: New approaches to congressional legislation and special interest groups* [Doctoral dissertation, New York University]. https://www.dropbox.com/s/b7ugpguf6fsufwm/Peterson_Dissertation.pdf?dl=0

Reh, C., Héritier, A., Bressanelli, E., & Koop, C. (2013). The informal politics of legislation: Explaining secluded decision making in the European Union. *Comparative Political Studies*, *46*(9), 1112–1142.

Ripoll Servent, A., & Panning, L. (2021). Engaging the dis-

engaged? Explaining the participation of Eurosceptic MEPs in trilogue negotiations. *Journal of European Public Policy*, *28*(1), 72–92.

Ripoll Servent, A. (2013). Holding the European Parliament responsible: Policy shift in the data retention directive from consultation to codecision. *Journal of European Public Policy*, *20*(7), 972–987.

Ripoll Servent, A. (2018). The European Parliament in justice and home affairs: becoming more realistic at the expense of human rights? In F. Trauner & A. Ripoll Servent (Eds.), *The Routledge handbook of justice and home affairs research* (pp. 385–395). Routledge.

Ripoll Servent, A. (2018a). *The European Parliament*. Palgrave Macmillan.

Roederer-Rynning, C., & Greenwood, J. (2015). The culture of trilogues. *Journal of European Public Policy*, *22*(8), 1148–1165.

Selck, T. J. (2006). *Preferences and procedures. European Union legislative decision-making*. Springer.

Senninger, R., Finke, D., & Blom-Hansen, J. (2020). Coordination inside government administrations: Lessons from the EU Commission. *Governance: An international journal of Policy, Administration, and Institution*. Advance online publication. https://doi.org/10.1111/gove.12525

Steunenberg, B. (1994). Decision-making under different institutional arrangements—Legislation by the European Community. *Journal of Institutional and Theoretical Economics*, *150*(4), 642–669.

Steunenberg, B. (2000). Constitutional change in the European Union: Parliament's impact on the reform of the codecision procedure. In H. Wagenaar (Ed.), *Government institutions: Effects, changes and normative Foundations* (pp. 89–108). Kluwer.

Thomson, R. (2011). *Resolving controversy in the European Union. Legislative decision-making before and after enlargement*. Cambridge University Press.

Thomson, R. (2015). The distribution of power. In J. Richardson & S. Mazey (Eds.), *European Union: Power and policy-making*. Routledge.

Thomson, R., & Hosli, M. (2006). Who has power in the EU? The Commission, Council and Parliament in legislative decision-making. *Journal of Common Market Studies*, *44*(2), 391–417.

Trauner, F., & Lavenex, S. (2015). A comparative view: Understanding and explaining policy change in the area of freedom, security and justice. In F. Trauner & A. Ripoll Servent (Eds.), *Policy change in the area of freedom, security and justice: How EU institutions matter* (pp. 219–240). Routledge.

Trauner, F., & Ripoll Servent, A. (Eds.). (2015). *Policy change in the area of freedom, security and justice: How EU institutions matter*. Routledge.

Tsebelis, G., & Garrett, G. (2000). Legislative politics in the European Union. *European Union Politics*, *1*(1), 9–36.

Tsebelis, G., Jensen, C. B., Kalandrakis, A., & Kreppel, A. (2001). Legislative procedures in the European Union: An empirical analysis. *British Journal of Political Science*, *31*(4), 573–599.

About the Author

Angela Tacea is a FWO (Research Foundation Flanders) Postdoctoral Fellow. She is affiliated with the Institute for European Studies of the Vrije Universiteit Brussel (VUB). She is also an associated researcher at the Center for Europeans studies and comparative politics (CEE), Sciences Po Paris. Her research and teaching interests are broadly in European decision-making process and constitutional law, with special focus on institutional actors and procedure, fundamental rights and justice and home affairs policies, e-government, civic techs, and artificial intelligence.