SSOAR

Open Access Repository
www.ssoar.info

Quantifying European Legislative Research: using CELEX and PreLex in EU legislative studies

König, Thomas; Luetgert, Brooke; Dannwolf, Tanja

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
König, T., Luetgert, B., & Dannwolf, T. (2006). Quantifying European Legislative Research: using CELEX and PreLex in EU legislative studies. *European Union Politics*, *7*(4), 553-574. https://doi.org/10.1177/1465116506069444

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-229118

European Union Politics
DOI: 10.1177/1465116506069444
Volume 7 (4): 553–574
Copyright© 2006
SAGE Publications
London, Thousand Oaks CA,
New Delhi

Forum Section

Quantifying European Legislative Research

Using CELEX and PreLex in EU Legislative Studies

Thomas König
German University of Administrative Sciences, Speyer, Germany

Brooke Luetgert
German University of Administrative Sciences, Speyer, Germany

Tanja Dannwolf
University of Mannheim, Germany

ABSTRACT

Research on European legislative decision-making has entered a stage of quantitative analysis. The quantitative approach promises to advance the current dialogue by allowing for the evaluation of competing approaches across multiple policy domains and over time. At the same time, the quantitative study of EU decision-making introduces a number of drawbacks: it is difficult to identify one definitive source for legislative information, and case-level data are not directly accessible in a machine-readable format. In order to identify the most crucial pitfalls and provide a reliable data source, we evaluate the most frequently cited, publicly available EU legislative database, CELEX, and compare it with a less publicized legislative database referred to as PreLex. We find that CELEX documents legislative events, whereas PreLex records inter-institutional activities in the legislative process. Unsurprisingly, each of these databases has particular advantages, and we discuss which of the two might be better suited for the analysis of specific research questions.

KEY WORDS

- CELEX
- European legislation
- legislative analysis
- PreLex

Legislative studies in the European Union

Research on legislative decision-making in the European Union (EU) has entered the stage of quantitative analysis, and this step promises to advance the current dialogue by evaluating and controlling for competing approaches (Jupille et al., 2003: 29).[1] After numerous case studies and intense theoretical debates, rich and systematic data analyses on the process of legislative decision-making may help in answering some of the most important questions raised in the controversies between rational choice analysts (Moravcsik, 1998; Schneider and Aspinwall, 2001) and constructivists (Checkel, 2001; Risse, 2000), between cooperative (Felsenthal and Machover, 2001; Hosli, 2000) and non-cooperative game theorists (Garrett and Tsebelis, 2001; Steunenberg et al., 1999; Tsebelis and Garrett, 2000), and among rational choice scholars themselves (Crombez, 1996, 2000; Moser, 1996; Steunenberg, 1994; Tsebelis, 1994, 2002). Has the EU moved towards majority voting, and to what extent does the European Parliament (EP) influence legislative decision-making? And what are the crucial factors influencing the process and outcomes – is the number of veto players decisive, or do we have to take other factors into account such as the kind of legislative instrument, the policy sector, the packaging or even the sequencing of proposals?

The legislative arena is the centre of European integration as the Commission, the member states and the EP adopt binding decisions that affect the lives of roughly 450 million citizens. Numerous legislative studies ask how EU institutions matter for these decisions, and whether institutional involvement varies across policy fields and procedures. In the past 15 years these institutional provisions have been modified several times, and the current European constitutional debate has centred around the distribution of voting weights, the involvement of the EP and access to the Commission in legislative decision-making (König and Hug, 2006). Quantitative analysis may shed light on the extent to which inter-institutional involvement and specific procedural rules affect legislative outcomes. We may look for changes in the application of particular procedures or voting rules in specific sectors and identify general trends in the quantity and type of legislative decisions over time, while considering the effects of treaty revisions and the accession of new members.

Although we possess a great deal of theoretical insight on these effects, the quantitative study of inter- and intra-institutional decision-making can provide cross-sectional and intertemporal evidence for many important claims about EU legislative processes. This evidence is not just required for EU legislative scholars. Implementation studies also suggest that the legislative process may influence correct and timely compliance with Community directives, and

thus, in the event of non-compliance, further infringement processes and decisions of the European Court of Justice (Giuliani, 2003; König et al., 2005b; Mastenbroek, 2003; Mbaye, 2001). Furthermore, the current debate on European constitution-building needs empirical support for normative claims about the impact of EU legislation on actors' constitutional preferences and the reform of the institutional framework (Tsebelis, 2005). For example, studies on public support assume that EU legislation 'hits home', even though the amount and extent of new EU legislative activities have significantly receded in recent years (Börzel and Risse, 2000; Gabel, 1998).

We do not believe that quantitative analyses will satisfactorily answer all of these questions or supersede further questions and debates on EU legislative decision-making. However, considering the large number of case-study insights, it is time for EU scholars to evaluate competing views from a more general perspective by including cases from multiple policy domains and over time (Franchino, 2000; Gabel et al., 2002). At the same time, we acknowledge that the quantitative study of EU legislative decision-making introduces a number of risks and pitfalls to which we would like to draw particular attention. In the few available empirical studies, we find, for example, very different reports of EU legislative activity.[2] For EU scholars interested in quantitative research, it is often difficult to identify one definitive source for this information, and case-level data are not easily accessible in a machine-readable format. Using different and sometimes incompatible sources of information may increase confusion rather than clarifying the different views on EU legislative research.

Our goal is to facilitate systematic research on EU legislative decision-making. To avoid the most crucial pitfalls and to provide a reliable source of quantitative information, we combine and evaluate the most frequently cited, publicly available EU legislative database, CELEX, and compare it with a less publicized but equally valuable legislative database known as PreLex. These data sources were developed independently by the Commission and a private provider, respectively, to fulfil different purposes; a fact that is also reflected in their unique structures.[3] Briefly stated, we find that CELEX documents legislative events, whereas PreLex records inter-institutional activities in the legislative process. Unsurprisingly, each of these databases – the event-oriented CELEX and the process-oriented PreLex – has particular advantages, and we would like to discuss which of the two might be better suited for the quantitative analysis of specific research questions. We also evaluate the reliability of these sources, although we would contend that most irregularities are user related.

This article presents the (dis-)similarities between CELEX and PreLex, describing their goals and distinctive composition. For the purpose of

comparison, we examine the period from 1984 to 2003, when the Nice Treaty came into force. Prior to this period, the official CELEX user's manual as well as the minimal number of documented initiatives and adoptions suggest that these databases are incomplete (König, 2001, 2006; König and Schulz, 1997; Schulz and König, 2000). Our censored CELEX data set includes 11,974 Commission proposals for binding legislation initiated between 1984 and 2003, while PreLex documents 9832 legislative processes for the same period. In the next section, we combine the process- and event-related information contained in both databases to cross-validate the documented cases and variables. Finally, we present the combined data set and describe the most important variables, such as the applied legislative procedure, the voting rule in the Council and the sector specification. The appendices document the technical transformation process from full-text internet resources to machine-readable data.[4]

PreLex – CELEX: Similarities and differences

In an effort to increase transparency in the EU legislative process, CELEX was launched in 1980 with the goal of making full-text documents publicly accessible. The database is divided into 12 sectors, each devoted to a different type of document. For example, sector 1 contains the EU treaties, sector 3 binding and non-binding legislation, sector 5 preparatory acts of all kinds, and sector 7 national implementation measures for Community directives. All sectors include detailed information on relevant dates, the legislative basis, the document title and related documents. In order to construct a data set containing all Commission proposals and the subsequently adopted legislation, we downloaded the Commission proposals in sector 5 (hereafter referred to as CELEX5) and binding secondary legislation – decisions, regulations and directives – in sector 3 (CELEX3).

PreLex systematically covers the period after 1976 and was designed to monitor inter-institutional procedures. According to the PreLex Information Guide, the database 'follows all Commission proposals (legislative and budgetary dossiers, conclusions of international agreements) and communications from their transmission to the Council or the EP'. The PreLex website provides a timeline for each Commission proposal, with the dates of different legislative and consultative stages, including references to all involved institutional actors. In addition to detailed information on the dates of transmission and adoption by different institutions, the user finds information on the applied legislative and voting procedures, the legal basis, responsible Directorates-General (DGs), links to all documents passed by the different institutions, dates of all readings, the request for and decision of any concil-

iation committee, and decisions made by the EP and the Commission at each stage in the legislative process. Furthermore, PreLex includes identification of each proposal as an A or B item in the Council, which indicates whether the act was adopted at the Council meeting without debate, after an agreement had already been reached within the Committee of Permanent Representatives (A Items), or was adopted with continued debate (B Items).[5]

At first glance, PreLex appears to contain more specific and policy-relevant information for EU legislative research. In fact, only PreLex documents the entire legislative process through all relevant stages and related institutions. CELEX, by contrast, seems to contain more detailed information on the treaty basis and titles of both the proposals and legislative acts, but the two CELEX data sets must be merged in order to identify the adoption date and transposition deadlines (where applicable) of Commission proposals. Although the merging of adopted legislation as contained in CELEX3 and Commission proposals documented in CELEX5 is possible, information about inter-institutional processes, such as the involvement of a conciliation committee in the co-decision procedure, is available only in PreLex. Furthermore, in CELEX we lack information on Commission proposals not identified in the CELEX3 data on adopted legislation (i.e. whether they were withdrawn or rejected). CELEX, however, also contains proposals from institutions other than the Commission as well as all binding and non-binding secondary legislation. These documents also include tertiary Commission legislation, which is frequently initiated through Commission White and Green papers. This suggests that using CELEX may be particularly advantageous for the analysis of legislation where no inter-institutional procedure (i.e. involvement of multiple institutions) is applicable.

For EU legislative research, we believe that a more detailed description of the two data sets is necessary. We have worked intensively with these data sets since the mid 1990s and have identified several advantages as well as a few perilous pitfalls in using them. Since most of the irregularities are user related, we hope that our experience will help EU scholars to conduct their analyses and test their claims. Note that both data sets are missing information for some important variables, such as the voting rule or the legislative procedure, which we may be able to supplement with information on the legal basis. However, in order to confirm the reliability of these data and our respective assumptions, we propose comparing these two independent online resources, which requires noteworthy technological expertise to transform them into a machine-readable data set. Whereas CELEX provides access to legal documents, PreLex reconstructs the inter-institutional legislative process; therefore, we must take different approaches in constructing comparable data sets.

CELEX

In order to transform the CELEX documentation into a machine-readable data set, we downloaded and combined information from sectors 5 and 3. Because different search commands and downloading dates can dramatically affect the sample size, we describe our downloading process in the appendix.

As previously mentioned, many variables of interest for EU legislative research are not documented in CELEX. The information on relevant dates, policy fields and cited legal basis is extracted directly. As indicated in Table 1, the document title clearly indicates the number of actors involved in passing the legislation (CELEX3) as well as the type of legislation. Note that, although our search was restricted to binding secondary legislation, we encountered a surprising number of corrigenda and tertiary Commission legislation that, by definition, are not found in the data on Commission proposals and amending drafts (CELEX5). Although the title of the adopted legislation (or draft proposal in the case of pending legislation) may be used to identify the application of co-decision, we are unable to distinguish among the procedures classified as Council legislation.

In order to infer the specific applied legislative procedure, we used the treaty basis. After generating a list of all treaty references mentioned in CELEX3 and CELEX5, we identified the legislative procedure associated with each legal base as recorded in sector 1.[6] We do not differentiate between the agreement and the consultation procedure in this analysis because both refer to Council legislation passed without EP participation. For cases in which several treaty articles were mentioned, or multiple procedures may apply, all procedures were coded, and we assume application of the strictest procedure,

Table 1 Summary of adopted legislation documented in CELEX3 (1984–2003)

	Type of legislation according to document title					
Document type	*Corrigenda*	*Commission*	*Council*	*Council & EP*	*EP/ ECB/other*	*Total*
Decision (D)	653	8666	2887	84	217	12,507
Directive (L)	903	648	879	225	2	2657
Regulation (R)	4475	17,135	6044	70	15	27,739
Total	6031	26,449	9810	379	234	42,903

Note: These data include binding secondary legislation passed between 1 January 1984 and 1 February 2003. Stricter censoring criteria will be applied in the following section. The distinction between document types is made according to the document code provided in the CELEX reference number: Decision (D), Directive (L), and Regulation (R).

where stricter is understood as more EP involvement. If only legislative acts were mentioned as legal bases, we infer the applicable legislative procedure from the cited legislation. The same assumptions apply to our coding of the voting rule applied in the Council; the strictest rule, or the rule presenting the highest voting hurdle, potentially prescribed in the treaty references is assumed.[7] Although information on the policy field is provided in CELEX, we also coded the legislative basis in order to obtain a classification according to the treaty chapters.

Several potential factors may contribute to missing values on these variables. First, only the legal references to standard EU treaties were coded. There are 367 cases where only an article of an accession treaty is mentioned. Second, in 214 cases no legal basis is mentioned. Finally, in 73 instances no legislative procedure could be identified for the provided legal reference (e.g. Article 130 of the Treaty establishing the European Economic Community refers to the European Central Bank). Ultimately, the applied legislative procedure and the Council voting rule are inferred for 11,308 legal acts.

Figure 1 shows the distribution of adopted legislation according to the year of its adoption. Most of the legislation recorded in the CELEX data set documents regulations. The absolute amount of legislation passed each year has declined rapidly since the Maastricht Treaty entered into force in 1993. The sharp decline after 2002 is largely attributable to pending legislation.

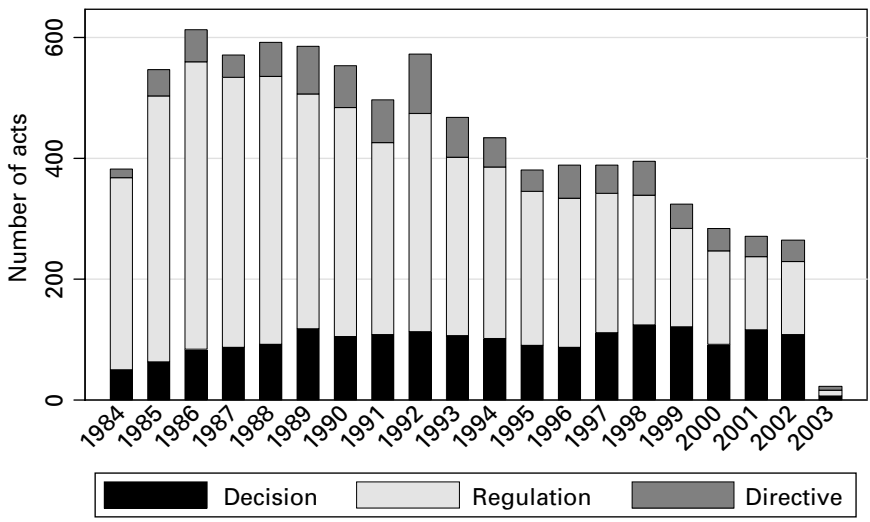

Figure 1 Distribution of adopted legislation in CELEX (1984–2003).
Notes: Illustration of data merged from CELEX5 and CELEX3 ($N = 11{,}974$). These data have been merged and censored as described in Appendix I. This figure illustrates the subsample of adopted legislation.

PreLex

In contrast to CELEX, PreLex was designed to document the entire legislative process of inter-institutional legislation. The richness of these data complicates the extraction of key characteristics, but also provides an invaluable insight into the complexity of Community legislative procedures. For the CELEX data, we merged Commission proposals with referenced passed legislation. This process leads to a rectangular data sheet documenting legislative events, where each proposal (including amended proposals) is assigned to one passed legislative act. Proposals lacking this reference remain in the data as pending legislation. In PreLex, at least seven unique legislative paths are documented. Table 2 reveals that, although over 82% of the legislation documented in PreLex involves one proposal being passed as one legislative act

Table 2 A typology of legislative processes identified in PreLex

	Type	N	*Details*	
Type 1:	One proposal One adoption	7547		
Type 2:	One proposal No adoption	646	235 405 6	pending withdrawn not adopted
Type 3:	One proposal & *n* amendments One adoption	1154	821 320 9 4	1 amendment 2 amendments 3 amendments 4 amendments
Type 4:	One proposal & *n* amendments No adoption	172	108 4 60 158 14	withdrawn not adopted pending 1 amendment 2 amendments
Type 5:	*n* proposals (some with amendments) One adoption	174	174 c. 82	proposals adopted acts
Type 6:	One proposal *n* adoptions	111	77 8 3 20 3	1:2 1:3 1:5 1:6 1:7–21
Type 7:	One proposal & *n* amendments *n* adoptions	28	20 2 2 2 1 1	2:2 2:3 3:2 3:3 4:2 4:3

(type 1) or one proposal still pending without reference to any amending proposals (type 2), the remaining legislation contains multiple proposals and/or multiple legislative acts.

For process-oriented legislative research, the PreLex data reveal how the strategic interaction of institutional actors may result in amended proposals (types 3, 4 and 7), package deals (type 5) or multiple legislative acts (types 6 and 7). Although this typology may not be exhaustive, and individual cases may change their type over time as existing proposals are re-examined or passed, this depiction illustrates how complicated legislative processes and related documents can be identified in a manner not possible in CELEX.

One potential disadvantage in PreLex is that the treaty references, or legal bases, are in a different and less detailed format. They are not as easily identifiable (lacking the link to the cited paragraph in the appropriate treaty version as found in CELEX sector 1) and are not as coherently coded as in CELEX, and the applicable treaty version is less clearly defined. Although small coding discrepancies make automated extraction more difficult, we transformed the identified legal bases into the CELEX format, where possible, and coded the legislative procedure applying the same method as described for the CELEX data set. The document number provided in PreLex is also in a different format but is easily transformed. Since the document numbers in both data sets appear to be based on the same general numbering scheme, we convert these numbers to the CELEX format and rely on them as unique case identifiers.

PreLex includes detailed information on the dates and decisions made in course of the legislative process, such as the decision of the EP after its first reading. A further set of variables is of special interest because this information can be obtained only indirectly in CELEX, or sometimes not at all. These variables are the documented legislative procedure, the voting rule in the Council, the responsible Directorate-General and the Council OJ item (the Council agenda item as documented in the *Official Journal*). As can be seen in Table 3, only 0.4% of all cases fail to mention information on the Council OJ item where we would expect it. Unfortunately, a much higher percentage of information on the procedure and voting rule variables is either not available or missing.[8]

Another defining characteristic concerns the responsibility of the Commission's DG. Table 4 shows the type and amount of adopted legislation for which selected DGs were responsible. If no adopted document is given, we differentiate between pending, withdrawn or rejected proposals and cases where an adopted document would be expected based on the events documented in PreLex (such as 'formal adoption by Council') but was not mentioned.

Table 3 Council OJ item, legislative procedure and Council decision mode: Tabulation of binding and censored PreLex data (1984–2003)

	No.	*%*
Last mentioned reference to OJ item		
OJ Council A Item	7556	76.9
OJ Council B Item	1111	11.3
Written procedure	363	3.7
Not available where expected[a]	41	0.4
Missing[b]	802	8.1
Total	9832	100.0
Legislative procedure[c]		
Co-decision procedure	440	4.5
Consultation procedure	2853	29.0
Cooperation procedure	544	5.5
Agreement procedure	780	7.9
Assent procedure	133	1.4
Consultation European Central Bank	6	0.1
Social protocol	3	0.0
Not available	5073	51.6
Total	9832	100.0
Council decision mode		
QMV	816	8.3
Unanimity	901	9.2
Written procedure	76	0.8
Not available	1339	13.6
Missing	6700	68.1
Total	9832	100.0

Notes:
[a] This category identifies the cases where this information should have been provided (based on the elected legislative procedure), but was not. For example, the formal adoption by the Council is documented but no information was provided on the relevant OJ Council item.
[b] For these cases, we would not have expected an OJ Council item to be mentioned, owing to the elected procedure; however, despite not having encountered such a case in repeated sampling, we cannot fully exclude the possibility of minor program errors.
[c] In addition to the common legislative procedures, PreLex also documents an 'agreement procedure'. This legislation is linked to proposals of decisions or regulations of the Council in the field of agreements and cites article 133 (Amsterdam consolidated version, previously article 113) as the legislative basis. In many cases an additional reference is made to article 300, paragraphs 1 and 2 (Amsterdam consolidated version). From these treaty references and the PreLex documentation, it may be assumed that the EP is not actively involved in the adoption of these proposals.

Although PreLex may be of particular interest for scholars studying the EU legislative process and inter-institutional interaction, we find many missing values on key variables: the legislative procedure and the voting rule.

Table 4 Distribution of legislation in selected DGs: Tabulation of binding legislation recorded in PreLex for the period 1984–2003

	Type of legislative act (adopted CELEX no. in header)						
Selected DGs	*Decision*	*Directive*	*Regulation*	*Pending*	*Withdrawn/ rejected*	*Not available*	*Total*
DG01 Trade	372	1	873	25	43	83	1397
DG03 Industry	33	206	73	3	20	11	346
DG05 Employment, Social Affairs	54	58	35	14	27	12	200
DG06 Agriculture & DG Fisheries	461	302	2112	24	161	43	3103
DG11 Environment	94	93	37	15	28	30	297
DG15 Internal Market	15	93	14	15	21	5	163
DG21 Taxation and Customs Union	240	43	854	26	53	41	1257
Others	899	172	1482	173	170	173	3069
Total	2168	968	5480	295	523	398	9832

Notes: Here we have selected eight of the most important DGs with respect to both the output and variance they represent. We combine the areas of fisheries and agriculture.

For policy studies, however, PreLex provides specific information on the primarily and jointly responsible DG. Agriculture, trade, taxation and customs union as well as internal market dominate the Community's legislative activities. As the previous results in CELEX suggest, most Commission proposals are passed as regulations. Decisions and directives are adopted less frequently, but this trend is also sector dependent, with notably more directives being passed in the areas of industry and the internal market.

Combining PreLex and CELEX: Reliability of data

Although the decision to use CELEX over PreLex, or vice versa, might be motivated by the research question at hand, the reliability of the provided indicators may be another concern. In order to compare these indicators in CELEX and PreLex, we merged both data sets using the initial proposal and the first adopted document as the unique merging criteria. Owing to the censored nature of the data, we exclude cases initiated prior to 1 January 1984 and cases passed or initiated after the coming into force of the Nice Treaty in February 2003. Figure 2 provides an overview of the different merge processes and the resulting number of cases. Unsurprisingly, not all legislation could be identified in both data sets. One explanation for this discrepancy rests in the event- vs. process-oriented designs of the respective databases. In the following, we concentrate on the successfully merged sample of 8475 cases and leave aside further speculation about these discrepancies.[9]

Figure 2 illustrates the cases we wish to compare. Recall that, for CELEX, we employed a two-stage process combining 13,001 Commission proposals from CELEX5 with the 42,903 cases of adopted legislation in CELEX3. PreLex, in contrast, documents 16,676 legislative processes, including non-binding legislation and Commission proposals lacking the standard PC prefix (e.g. those based on so-called SEC documents, including White and Green papers). Censoring both data sets to include only binding legislation for the period under study reduces the samples to 11,974 and 9832 cases, respectively. Merging both data sets further reduces the sample to 8475 cases, because 3499 cases from CELEX and 1357 cases from PreLex could not be identified in the other data set. This does not necessarily suggest that CELEX documents a more complete sample of Commission proposals than PreLex. The difference in numbers is primarily attributed to amending proposals, which are documented as separate cases in CELEX but are counted as one legislative process in PreLex, and which are therefore not included in the merged data set.

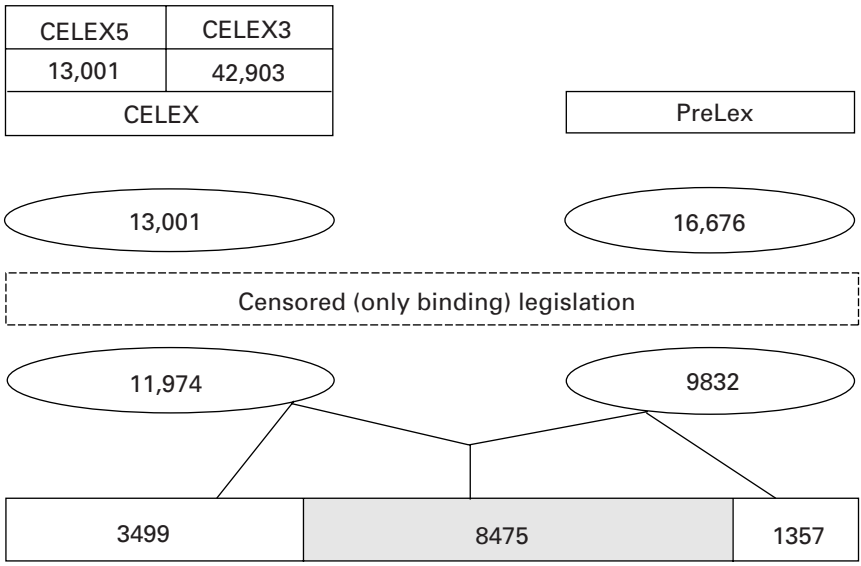

Figure 2 Overview of the data sets.

Legislative procedure

Institutional analyses frequently focus on the impact of the legislative procedures. Whereas the legislative procedure must be coded in CELEX, this information is partially documented in PreLex. This is a valuable means of testing reliability, even though we find more than 50% missing values. According to the legal bases cited for the cases lacking a documented legislative procedure in PreLex, it is reasonable to assume no EP involvement and that these cases were adopted by the Council under the qualified majority (QMV) rule. Table 5 provides a comparison of the coded legislative procedure obtained in CELEX and the information in PreLex.

According to Table 5, reliability is limited because information for a number of cases is not available or there are coding discrepancies. This may be a result of our choice to apply the strictest criterion for coding parliamentary participation when multiple treaty articles were indicated in CELEX. To confirm reliability, we compare the values obtained by extrapolating the procedure from the legal basis in both data sets. Table 6 shows that we obtain similar results. This suggests that scholars should pay careful attention to the drawbacks of procedural indicators (e.g. missing values, assumptions, etc.).

To sum up, information on the legislative procedures is best obtained from various sources. The coding of the legal basis leads to the least number

Table 5 Cross-tabulation of legislative procedure

	Role of EP (CELEX)				
Legislative procedure (PreLex)	*Consultation/ no EP involvement*	*Cooperation*	*Co-decision*	*Not identified*	*Total*
Consultation procedure	**2194**	32	221	76	2523
Cooperation procedure	249	**167**	59	21	496
Co-decision procedure	122	9	**220**	3	354
Agreement	502	1	49	0	552
Assent procedure	35	1	57	0	93
Consultation ECB/Social protocol	6	0	0	0	6
Not available	3865	32	70	484	4451
Total	6973	242	676	584	8475

Notes: Tabulation of successfully merged binding and censored PreLex and CELEX data (1984–2003). EP involvement is coded according to the mentioned treaty basis in CELEX and cross-tabulated with the procedure documented in PreLex. We again include the agreement procedure as documented in PreLex (refer to Table 3).

Table 6 Legislative procedure inferred using the cited legal basis: A comparison of PreLex and CELEX

	Role of EP according to treaty basis (CELEX)				
Role of EP according to treaty basis (PreLex)	*Consultation/ no involvement*	*Cooperation*	*Co-decision*	*Not identified*	*Total*
Consultation/no involvement	**4194**	66	259	27	4546
Cooperation	230	**134**	45	18	427
Co-decision	83	3	**298**	1	385
Not identified	2466	39	74	538	3117
Total	6973	242	676	584	8475

of missing values, but cross-validation with documented procedures in PreLex reveals some discrepancies. The application of co-decision is easily determined for adopted legislation according to the document title in CELEX; however, when more detailed information is required (e.g. application of consultation or cooperation procedures, particularly for pending proposals), there may be no alternative to manually coding the legal basis.

Voting rule in the Council

In addition to the legislative procedure, other studies also focus on the applied Council voting rule. Table 7 shows the cross-tabulation of the indicator in PreLex and the extrapolated reference to the voting rule in CELEX. The results reveal an underestimation of the application of unanimity: 55% of the cases documented in PreLex as unanimity are classified as QMV in CELEX. Owing to the phrasing of the treaties, the voting rule is more difficult to deduce from the legal basis than is the legislative procedure. As with the legislative procedure, the high number of missing values compels both PreLex and CELEX users to code the legal basis manually if information is desired on the voting rule applied in the Council.

A comparison of the voting rule inferred from the respective legal basis in both data sets again shows that similar results are obtained (see Table 8). Compared with CELEX, PreLex reveals a more consensus-oriented bias towards the unanimity rule but also contains many missing values.

Policy field

Policy studies frequently refer to the specific feature of a policy area of EU legislative decision-making. A comparison of the listed and/or extrapolated policy fields and the primarily responsible DGs may provide some cross-validation. However, information on the degree to which DGs are responsible for legislation in given policy fields is limited. In PreLex, the only information

Table 7 Voting rule in Council: A comparison of PreLex and CELEX

	Voting rule according to treaty basis (CELEX)			
Last reference documented (PreLex)	*Unanimity*	*QMV*	*Not identified*	*Total*
Unanimity	**325**	469	51	845
QMV	148	**587**	41	776
Written procedure	6	0	0	6
Not mentioned	469	537	60	1066
Missing	1709	3641	432	5782
Total	2657	5234	584	8475

Notes: The legal basis in PreLex is not substituted if a legal act is cited. In CELEX, we have substituted these references. After repeated sampling of the cited legal bases we note, however, that these cases are likely Council legislation approved under QMV and not subject to inter-institutional procedures.

Table 8 Voting rule in Council inferred using the cited legal basis: A comparison of PreLex and CELEX

	Voting rule according to treaty basis (CELEX)			
Voting rule according to treaty basis (PreLex)	*Unanimity*	*QMV*	*Not identified*	*Total*
Unanimity	**1161**	68	23	1252
QMV	820	**3263**	23	4106
Not identified	676	1903	538	3117
Total	2657	5234	584	8475

provided on the related policy sector is the primarily and jointly responsible DG. In CELEX, the policy sector may be obtained by coding the legal basis according to the treaty chapter sub-heading or from the EUROVOC classification provided by CELEX. For a better overview, we have classified this policy sector information into five categories. Table 9 shows the distribution of the legislation in the eight most common DGs according to the classified policy sector.

For research on the legislative process conducted using CELEX, it is valuable to know that the responsible DG can be deduced relatively accurately from the policy field information. This is especially true in the area of agriculture, but may be more problematic in areas related to the common market and industry.

Table 9 indicates that the cases in the most important policy areas largely coincide. This does not mean that overlapping policy competencies result only from coding. Rather, this suggests that many proposals involve multiple competencies – a well-known problem within the Commission. For this reason, the preparation of many proposals includes several DGs.

Conclusion

Our discussion of the data sets has shown that there are pros and cons with both CELEX and PreLex. Using the PreLex data, we find that roughly 17% of EU legislation cannot be easily classified in a rectangular data sheet relating one proposal to one passed legislative act. Complicated legislative processes are better documented in PreLex because related amending proposals and resulting multiple acts are connected in individual cases. If, however, we want to trace legislative events and consider amending proposals as individual

Table 9 Primarily responsible DG according to the categorized policy field

	Selected DG									
CELEX policy field	*DG01 Trade*	*DG03 Industry*	*DG05 Employment, Social Affairs*	*DG06 Agriculture & DG Fisheries*	*DG11 Environment*	*DG15 Internal Market*	*DG21 Taxation and Customs Union*	*Other*	*N/A*	*Total*
Trade	844	62	1	337	23	2	522	349	1	2141
Common rules	3	200	18	23	57	43	203	125	0	672
Agriculture	174	19	6	1923	2	0	5	851	2	2982
Internal market	14	20	26	11	0	60	259	144	0	534
Various	136	20	119	600	174	32	98	963	4	2146
Total	1171	321	170	2894	256	137	1087	2432	7	8475

cases, CELEX may provide the more attractive structure. Either way, we have established that many of the contradictory claims about EU legislative activity may be reducible to the question of whether events or processes are counted. Furthermore, we observe differences in the period and types of document recorded in these databases.

In our view, CELEX is an integrated, highly specialized and differentiated collection of legislative documents. The direct link to the full-text documents, the treaty citation and the national implementation measures are advantageous for analysing EU legislative acts. The data also include Commission tertiary legislation as well as corrigenda and non-binding documents. This is particularly relevant for implementation studies and those analyses examining the possible impact of norms and informal rules imposed by non-binding acts. However, the need to link documents and merge characteristics to document legislative events may provide an additional source of error, if not approached systematically. The tabular structure is more straightforward in its construction and its citation of the legal treaty basis, but the need to code the legislative bases is a disadvantage that also persists in PreLex.

Researchers seeking detailed information on the inter-institutional legislative process may find PreLex more suited to their needs. Here, we have not yet explored the information on the available dates and decisions made at each reading, but this might prove to be a valuable task. Particularly for those interested in the involvement of the EP, PreLex provides references to parliamentary documents, rapporteurs and transmission dates. The unique advantage of PreLex is that the entire inter-institutional process is documented to include all actors involved and their decisions at each legislative stage.

Although it seems plausible to combine information from both data sets, it should be kept in mind that some discrepancies remain, which may result from the different purposes of these resources. As previously mentioned, the document numbering (particularly after the hyphen) differs in some cases. Combining the two data sets may lead to specific biases; however, missing values can be minimized, the quality of some variables such as the voting rule and the legislative procedure can be improved, and additional information such as the responsible DG or related Council OJ item may be supplemented. We recommend using the merged data set, but we feel that the dangers of merging data must be carefully considered, particularly for cases involving amended proposals and/or proposals leading to multiple resulting acts.

To promote research we have released the merged data set, although we believe that all data sets presented are valuable assets for future quantitative

activities. We hope that these data will help to resolve conflicts between competing theoretical strands and aid qualitative researchers in selecting representative legislation for their case studies. Our experience of more than 10 years of database management suggests that researchers usually make a number of explicit and implicit decisions when extracting and coding the full-text information. Sometimes, missing values occur because of a mishandling of the full-text information (see, for a discussion, König et al., 2005a). Avoiding these pitfalls is a very time-consuming process, and perhaps our release of these data will help other scholars to save time and run their analyses on a common and transparent ground. It is our belief that using reliable data will support progress in EU legislative research, whether these advances are used to develop new and better theories, to test existent theories, or to improve legislative decision-making.

Notes

1 The merged data set presented in this contribution may be downloaded at: http://www.dhv-speyer.de/tkoenig/.

2 Alesina et al. (2005: 291) used a CELEX search and identified 25,472 directives, regulations and decisions for the period 1986–1995. By contrast, König (2001) refers to 5701 Commission proposals for binding secondary legislation documented in CELEX for the period 1984–95. Although other studies also cite legislative statistics according to CELEX, these studies illustrate the wide range of numbers cited by researchers using the same data source. Without easily accessible official statistics on EU legislative activity, we must ask how these differing statistics come about. In this example, we predict that Alesina et al. (2005) cite sector 3 CELEX data and that their numbers may include corrigenda, Commission tertiary legislation and (possibly) duplicate cases, whereas König (2001, 2006) as well as König and Schulz (1997) and Schulz and König (2000) document Commission proposals.

3 CELEX, which stands for Communitatis Europae Lex, was updated on a daily basis until January 2005, when it was integrated into the EUR-Lex framework; however, for the period under study CELEX is still freely accessible at http://europa.eu.int/celex/. PreLex is also accessible via EUR-Lex and linked to the official European Union website at http://europa.eu.int/prelex/apcnet.cfm?CL=en.

4 The appendices are available at http://www.uni-konstanz.de/eup/issues.htm.

5 Although reference to the Council OJ items may not be the best indicator of conflict or of divergent member state preferences, it is the only indicator documented on a case-wise basis in these data.

6 The treaty references are recorded as CELEX sector 1 document numbers and identify individual treaty articles. In some instances reference is made to a particular sentence within an article, but in such cases the associated legislative process is not always easily identifiable.

7 Some may question our assumption regarding application of the strictest procedure mentioned; however, given a list of multiple legal bases, a rule of thumb must be defined. The appropriateness of this assumption is explored in the merged data set. Based on Tables 6 and 7, we do not find evidence of an overestimation of the application of unanimity.

8 Owing to the nature of the PreLex data, we suspect that these missing values have a systematic explanation. Repeated sampling of the legal bases cited in these cases confirms that this category largely refers to Council legislation adopted under QMV.

9 It should be noted that several documents share the same 11-digit document number, differing only in the sub-numbering after the hyphen, and that the documents associated with these sub-numbers differ. After merging the CELEX and PreLex data, we noticed that numbering schemes for proposals after the hyphen are not identical in the two databases. Therefore, the merging of these proposals was first done according to the 11 leading digits and confirmed using the adopted document number. A comparison of dates and document titles confirms this method. Unfortunately, proposals lacking an adopted document could not be merged owing to the lack of a unique identifier.

References

Alesina, Alberto, Ignazio Angeloni and Ludger Schuknecht (2005) 'What Does the European Union Do?', *Public Choice* 123: 275–319.

Börzel, Tanja and Thomas Risse (2000) 'When Europe Hits Home: Europeanization and Domestic Change, EUI Working Papers.' RSC No. 2000/56.

Checkel, Jeffery (2001) 'Constructing European Institutions', in Gerald Schneider and Mark Aspinwall (eds) *The Rules of Integration: Institutionalist Approaches to the Study of Europe*, pp. 19–40. Manchester: Manchester University Press.

Crombez, Christophe (1996) 'Legislative Procedures in the European Community', *British Journal of Political Science* 26: 199–228.

Crombez, Christophe (2000) 'Institutional Reform and Codecision in the European Union', *Constitutional Political Economy* 11: 41–57.

Felsenthal, Dan S. and Machover, Moshé (2001) 'Myths and Meanings of Voting Power', *Journal of Theoretical Politics* 13(1): 81–97.

Franchino, Fabio (2000) 'Control of the Commission's Executive Functions: Uncertainty, Conflict and Decision Rules', *European Union Politics* 1(1): 63–92.

Gabel, Matthew (1998) *Interests and Integration: Market Liberalization, Public Opinion, and European Union*. Ann Arbor: University of Michigan Press.

Gabel, Matthew, Simon Hix and Gerald Schneider (2002) 'Who Is Afraid of Cumulative Research? Improving Data on EU Politics', *European Union Politics* 3(4): 481–500.

Garrett, Geoffrey and George Tsebelis (2001) 'Even More Reasons to Resist the Temptation of Power Indices in the EU', *Journal of Theoretical Politics* 13(1): 99–105.

Giuliani, Marco (2003) 'Europeanization in Comparative Perspective: Institutional

Fit and National Adaptation', in Kevin Featherstone and Claudio Radaelli (eds) *The Politics of Europeanization*, pp. 134–55. Oxford: Oxford University Press.
Hosli, Madeleine (2000) 'The Creation of the European Economic and Monetary Union', *Journal of European Public Policy* 7(5): 744–66.
Jupille, Joseph, James Caporaso and Jeffery Checkel (2003) 'Integrating Institutions: Rationalism, Constructivism, and the Study of the European Union', *Comparative Political Studies* 36: 7–40.
König, Thomas (2001) 'Pivotal Politics in European Legislation: An Event History Analysis of the EU Legislative Decision-making Process from 1984 to 1995', paper delivered to the 2001 ECPR Joint Sessions.
König, Thomas (2006) 'Convergence or Divergence? From Ever-Growing to Ever-Slowing European Decision-making Process', *European Journal of Political Research*, forthcoming.
König, Thomas and Simon Hug (2006) *Policy-making Processes and the European Constitution: A Comparative Study of Member States and Accession Countries*. London: Routledge.
König, Thomas and Heiner Schulz (1997) 'The Efficiency of Legislative Decision Making in the European Union: An Event-History-Analysis of EU Decision Making Speed', Center for German and European Studies, University of California, Berkeley Working Paper 1.54.
König, Thomas, Daniel Finke and Stephanie Daimer (2005a) 'Ignoring the Non-ignorables? Missingness and Missing Positions', *European Union Politics* 6(3): 269–90.
König, Thomas, Brooke Luetgert and Lars Mäder (2005b) 'Troubles with Timeliness: Explaining Trends in Transposition Delay', paper delivered to the 2005 ECPR Joint Sessions.
Mastenbroek, Ellen (2003) 'Surviving the Deadline: Transposition of EU Directives in the Netherlands', *European Union Politics* 4(4): 371–95.
Mbaye, Heather (2001) 'Why National States Comply with Supranational Law', *European Union Politics* 2(3): 259–81.
Moravcsik, Andrew (1998) *The Choice for Europe: Social Purpose and State Power from Messina to Maastricht*. Ithaca, NY: Cornell University Press.
Moser, Peter (1996) 'The European Parliament as a Conditional Agenda Setter: What Are the Conditions?', *American Political Science Review* 90: 834–8.
Risse, Thomas (2000) 'Let's Argue!: Communicative Action in World Politics', *International Organization* 54: 1–39.
Schneider, Gerald and Mark Aspinwall (eds) (2001) *The Rules of Integration: The Institutionalist Approach to European Studies*. Manchester: Manchester University Press.
Schulz, Heiner and Thomas König (2000) 'Institutional Reform and Decision-Making: Efficiency in the European Union', *American Journal of Political Science* 44(4): 653–66.
Steunenberg, Bernard (1994) 'Decision Making under Different Institutional Arrangements: Legislation by the European Community', *Journal of Theoretical and Institutional Economics* 150: 642–69.
Steunenberg, Bernhard, Dieter Schmidtchen and Christian Koboldt (1999) 'Strategic Power in the European Union: Evaluating the Distribution of Power in Policy Games', *Journal of Theoretical Politics* 11(3): 339–66.

Tsebelis, George (1994) 'The Power of the European Parliament as a Conditional Agenda Setter', *American Political Science Review* 88: 128–42.
Tsebelis, George (2002) *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.
Tsebelis, George (2005) 'Assessing the Contributions of the DOSEI Project', *European Union Politics* 6(3): 377–90.
Tsebelis, George and Geoffrey Garrett (2000) 'Legislative Politics in the European Union', *European Union Politics* 1: 5–32.

About the authors

Thomas König is Professor of Political Science, German University of Administrative Sciences, PO Box 1409, 67324 Speyer, Germany.
Fax: +49 6232 654127
E-mail: tkoenig@dhv-speyer.de

Brooke Luetgert is a research assistant at the German University of Administrative Sciences, PO Box 1409, 67324 Speyer, Germany.
Fax: +49 6232 654127
E-mail: luetgert@dhv-speyer.de

Tanja Dannwolf is a graduate student at the University of Mannheim, Germany.
E-mail: tdannwol@rumms.uni-mannheim.de