

Appendix I: Construction of Machine-Readable CELEX Data

Based on a method developed in earlier studies (König and Schulz 1997; Schulz and König 2000; König 2001, 2006), our current efforts expand the period of analysis and apply a uniform procedure to the entire period of study in order to capture additional and/or corrected information in earlier legislation. Commission proposals documented in CELEX Sector 5 were last downloaded between 4 March 2004 and 26 April 2004 for the period from 1 January 1980 until 1 February 2003. Using the 'Expert Search' mode, we searched by year using the commands “Date_of_document >= YYYY/01/01 <= YYYY/12/31 AND Type_Type = PC”. The search results were displayed in tabular format and saved as ASCII text files. The same method was used to download information on binding secondary legislation using the search command “Type_Sector = 3 AND Type_Type = (D OR L OR R) AND Date_of_publication >= YYYY/01/01 <= YYYY/12/31”. These files were downloaded between 7 April and 19 May 2004 and cover the period from 1 January 1980 until 1 February 2003 (entry into force of Nice Treaty).

Using a special software developed for this purpose, we extracted multiple characteristics from the legislation documented in the ASCII files and imported these identifying features into a ProCite database.¹ Within ProCite, we confirmed the correct extraction of the desired characteristics for our subsequent analysis. In a second step, we imported the ProCite data into STATA 8.0 for coding and analytical purposes. We began with the CELEX5 data and merged information on adopted documents from CELEX3 in accordance with the adopted document mentioned in CELEX5. In CELEX, information on the legislative procedure is coded for each proposal according to the cited legal basis. CELEX5 contains 12,655 Commission proposals. For packaged proposals (several legislative acts adopted on the basis of one proposal), we multiplied the cases in the dataset meaning that

each legislative act is associated with identical proposal characteristics. This slightly increases the sample to a total of 13,001 cases, of which 9144 mention an adopted document.

The dataset is censored for the period from 1 January 1984 to 1 February 2003 using the proposal and adoption date. The censored dataset contains 12,484 proposals, of which 8891 mention an adopted document. Since our focus is on binding legislation, the dataset is further reduced to 11,974 cases, including 8764 pieces of adopted legislation. Non-binding legislation is identified by the type of the adopted document (where available) and the type of proposed legislation when no adopted document is available. CELEX3 includes a total of 51,151 binding legislative acts. This is reduced to 42,903 acts for the period from 1 January 1984 to 1 February 2003. Due to the inclusion of significantly more documents in CELEX3, we identify corrigenda and Commission tertiary legislation using the title of CELEX3 documents before merging adopted legislation to the Commission proposals included in CELEX5. The censored CELEX3 dataset contains 6031 corrigenda and 26,449 pieces of Commission tertiary legislation, leaving 10,189 legislative acts that could be referred to as adopted legislation in CELEX5. Depending on the research question at hand, we note that including these data in statistics on the legislative activity of the EU may be misleading.

In order to create a dataset allowing for research on the European legislative process, both CELEX datasets were merged using the adopted legislative document as the decisive merge criterion. With the exception of 171 cases, all binding legislation referred to in CELEX5 was found in CELEX3. Most of the documents not identified in the merge process were not part of the downloaded period of CELEX3 documents.² Thus, the merged CELEX data document 11,974 legislative events.

Appendix II: Construction of Machine-Readable PreLex Data

PreLex provides key identifying characteristics for Commission proposals as well as all related inter-institutional processes and adopted documents in HTML-format on the official website of the European Union (<http://europa.eu.int/prelex/apcnet.cfm=en#>). In an automated process on 11 January 2004, each documented Commission proposal (COM document) was downloaded and saved in a unique file. A specially developed software was used to extract the desired characteristics (date of initiation, date of vote in the European Parliament, date of vote in the Council, date of passing, applied procedure, adopted legislation etc.) from the HTML files and create a machine-readable STATA data sheet. Although PreLex documents selected cases from as early as 1967, prior to 1980 we found fewer than 35 cases documented annually. Hence, the downloaded 16,676 cases cover the period from 1980 to 2004 and include all adopted (binding as well as non-binding) legislation.

PreLex documents the entire legislative process, and because of this we do not have to perform any additional merging with other data. Furthermore, these data do not list individual amending proposals as unique cases, which reduces our sample to legislative processes identified by the first proposal.

The COM (or Commission) documents downloaded in PreLex include more than just Commission proposals. Furthermore, the sample documents a longer period than CELEX. Due to these differences, we censored the PreLex data to cover the same period and include only binding legislation based on Commission proposals. 2974 cases do not fall under the period of observation, and 3870 additional cases referred to non-binding documents. Thus, the dataset reduces to 9832 Commission proposals.

¹ Unlike other spreadsheet software, ProCite has the advantage that each field may include more than the standard 256 characters. This allows us to retain the full-text of the title and other longer descriptors.

² According to the year mentioned in the CELEX document number, 111 documents were adopted in 2003 or 2004. An additional 60 binding legislative documents should have been in our CELEX3-dataset but were not found. There is one series of 14 missing directives in 2001 having numbers between 301L0095 and 301L0115. Taking these facts into consideration, these missing data appear negligible.