CEPS EurLex dataset – Codebook Summary

This codebook provides an overview of the different variables available in the CEPS EurLex dataset[1]. Every datapoint was scraped from the official eur-lex.eu website with the programming languages R and Python. The variables therefore represent the information as it is available on the website without additional interpretation by our researchers. To better understand the structure of the data source and how the data is represented in its raw html format, please consult, for example, the eur-lex.eu website for the General Data Protection Regulation: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679

In essence, our scraper simply extracted relevant information from the website and transformed it into a CSV file, with every row representing one law and every column representing the different variables listed below. Please note that not all variables are available because the Eur-lex.eu website does not consistently provide data for all the variables. In addition, the HTML documents were not always cleanly formatted and text extraction from scanned PDFs is not entirely clean. Some data points are therefore missing for some laws and some laws were excluded entirely. In addition, not not all (older) laws were available in English, especially since Ireland and the UK only joined the European Communities in 1973. Non-English laws are excluded from the dataset.

The table provides an overview of: (1) the variable name as provided in the dataset; (2) a short explanation of the variable; (3) an example value from one cell to illustrate the format; (4) the number of legal acts containing a value for the specific variable.

Table 1 - Overview: Variables available in the CEPS EurLex dataset

Variable Name	Explanation	Example value	Number of legal acts
Act_amends	CELEX number of the old act amended by the new act (see details on CELEX below)	32013R1311	36.307
Act_cites	CELEX number of other acts cited by the act	32013R1311; 32018D508	65.647
Act_name	Full name of act	Decision (EU) 2018/846 of the European Parliament and of the Council of 30 May 2018 on the mobilisation of the …	139.889
act_raw_text	The full raw text of the act in one string. Mostly includes: title, recitals, legal articles and annex. Please note that the text of older laws is not always clean.	Long string of text, e.g. "DECISION (EU) 2015/2240 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 November 2015 ...."	134.633
Additional_info	Additional information	"EEA relevance", or "Validity: end of financial year 2016"	28.686
Ammends_links	Link to previous act which is amended by the new act	http://data.europa.eu/eli/reg/1968/823/oj	36.307
Authors	Name of the act's authors	European Parliament; European Council	142.035

CELEX	Unique CELEX identifier of the act. For details, see https://eur-lex.europa.eu/content/help/faq/celex-number.html	32019D0276	142.035
Cites_links	Link to other acts, cited by the act	http://data.europa.eu/eli/treaty/teec/art_233/sign; http://data.europa.eu/eli/reg/1964/111/oj	65.647
Date_document	Date of the document. The eur-lex.eu website does not provide an explanation of which exact date in the legislative process this represents. The dataset ranges from 1952 to August 2019.	2018-12-12	142.035
Date_publication	Date the document was published	2019-02-22	39.868
ELI_link	European Legislation Identifier (ELI) link to the act. (Mostly leads to same content as Eurlex_link)	http://data.europa.eu/eli/dec/2018/1675/oj	142.035
EUROVOC	A group of EuroVoc keywords associated with the act. For details see https://op.europa.eu/en/web/eu-vocabularies	ENP; economic stabilisation; change of economic system; EU macro-financial assistance; Ukraine; IMF; structural adjustment	91.613
Eurlex_link	Link to act on eur-lex.europa.eu website. (Mostly leads to same content as ELI_link)	https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32018D0846	142.036
First_entry_into_force	Date when act first entered into force	2019-01-01	104.135
Legal_basis_celex	The CELEX number of the act's legal basis	32013Q1220(01)	109.231
Oeil_link	Link to the European Parliament's Legislative Observatory (Oeil). Provides procedural information	http://www.europarl.europa.eu/oeil/FindByProcnum.do?lang=en&procnum=COD/2016/0039	5.263
Procedure_number	Number of the legislative procedure leading to the act	2018/0058/COD	18.622
Proposal_link	Link to the Commission proposal preceeding the act	https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52018PC0150	14.807
Status	Whether the act was in force at the time of scraping (August 2019). ("In Force" or "Not in Force")	In Force	85.312
Subject_matter	Group of keywords representing the subject matter of the act. Similar to EUROVOC, only less detailed, more abstract.	cooperation policy; budget; EU finance; international security; migration	91.613
Temporal_status	Date of end of validity of the act	2021-12-31	18.479
Treaty	Name of the Treaty the act is based on	TFEU	109.648