Codebook: EU Social Policy database

Daniel Fernandes, Leiden University

November 30, 2024

The EU Social Policy database was submitted as Deliverable 8.4 for the Horizon Europe research project *TransEuroWorkS: Transforming European Work and Social Protection*. This deliverable aims to compile harmonised and systematic data on EU-level social policy regulation.

1 Description **1**

2 Variables **3**
2.1 Adopted acts 3
2.2 Procedures 5
2.3 European Parliament resolutions 7
2.4 Transpositions of adopted acts to national law 9
2.5 Voting records 11
2.6 Explanations of votes 12
2.7 Information about members of Parliament 15

3 Linking datasets **18**
3.1 Direct links 18
3.2 Other combinations 19

1 Description

Objective: The EU Social Policy database (EUSPD) is a series of datasets providing data on policy-making at the EU level in the domains of social policy, education, and training.

Included data: The EUSPD comprises seven datasets. These include details on (1) procedures and (2) adopted acts classified in the Eur-Lex (2024) portal as either social policy or education and training. They also contain information on the legislative process, focusing on the European Parliament:

(3) parliamentary resolutions, (4) voting records, (5) explanations of votes, and (6) data on Members of the European Parliament (MEPs). Finally, the EUSPD includes all recorded (7) transpositions of adopted acts into national law. Section 2 describes the variables included in each dataset.

Links: These seven datasets are interlinkable. Section 2 identifies the linking variables, while Section 3 provides documentation on combining these datasets using the linking variables.

Data gathering process: The following process was used to compile the EUSPD. First, the export tool of Eur-Lex (2024) was used to download the base datasets containing document-based data: adopted acts, procedures, European Parliament resolutions, and transpositions of adopted acts to national law. Documents were selected based on directory codes 0520 (social policy) and 1630 (education and training).

Unique CELEX identifiers were used to link these documents. Although Eur-Lex (2024) provides extensive information about relationships between policy-making stages, its data are incomplete. The EUSPD supplements this information using data from the European Parliament, sourced from Parltrack (2024) and the Legislative Observatory of the European Parliament (2024). Additionally, the Golub (2024) EUPROPS dataset was used to complete adopted acts $\rightarrow$ procedure links and to ensure data quality.

Once the document-based data were compiled, the next step focused on European Parliament data. The Legislative Observatory of the European Parliament (2024) website was scraped for voting records and explanations of votes on EP resolutions included in the dataset. Details about Members of Parliament (MEPs) were also gathered for all mandates with available voting information.

MEPs' names were used to link these datasets. Unlike document-based data, no missing data were observed in these datasets. However, names were not standardised. Original MEP data formatted names as `name(s) SURNAME(S)` in Latin script. Explanations of votes followed the same structure but used original language scripts – Cyrillic for Bulgaria, Greek for Cyprus and Greece, and Latin for other languages. Voting records included only surname(s), except in cases where multiple individuals shared the same surname.

To match these datasets, names were standardised. Diacritics were removed, all text was converted to lowercase, and Cyrillic and Greek scripts were transliterated to Latin using the `cyrtranslit` Python package. In case the explanation of vote was signed by multiple MEPs only the first author was considered. This is a coding choice explicitly made to avoid duplicating explanation texts (assigned to each politician individually) in the final dataset.

Using these standardised names, we performed *step-wise matching*. This involved creating a merge variable based on various matching conditions, such as exact matches by surname-term-party, surname-term (for MEPs who changed party affiliation during a mandate), partial matches of full names, and manual coding for remaining observations. For each combination, duplicates in either the *master* or *using* dataset were removed to ensure unambiguous identification of individuals.

All import scripts and web scraping were conducted in Python using the `requests` and `BeautifulSoup` libraries. The matching process was implemented in Stata. To access the original code files, please email the author of EUSPD at transeuroworks@law.leidenuniv.nl.

2 Variables

2.1 Adopted acts

File name: `adopted_acts.dta`
Description: This dataset contains all EU adopted acts related to social policy, education and training.
Unit of observation: Adopted act.

Variable name:	`act_name`
Description:	Title of the adopted act.
Type:	string
Example:	Council Recommendation of 27 November 2023 on developing social economy framework conditions.

Variable name:	`act_number`
Description:	CELEX number of the adopted act. Linking variable
Type:	string
Example:	32023H01344
Note:	CELEX numbers are unique identifiers used for European Union legal documents. They have four parts: the legal sector, the year of publication, the category of legal act (e.g., directive, regulation), and a unique document number. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`procedure_number`
Description:	Procedure number of the adopted act. Linking variable
Type:	string
Example:	2023/0179/NLE
Note:	Procedure numbers have three parts: the year, the procedure number (digits and/or letters) and abbreviation of the procedure type. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website. Adopted acts originating from processes internal to specific European institutions or bodies are excluded from this variable. In such cases, this variable is coded as missing.

Variable name:	`act_type`
Description:	Type of adopted act.
Type:	numeric (labeled)
Label list	1 Budget 2 Communication 3 Consolidated text 4 Council conclusions 5 Decision 6 Declaration 7 Directive 8 General guidelines 9 Implementing decision 10 Joint action 11 List 12 Opinion 13 Programme 14 Recommendation 15 Regulation 16 Report 17 Resolution 18 Rules of procedure 19 Special report 20 Statement 21 Statute

Variable name:	`act_eurovoc`
Description:	EuroVoc keywords of the adopted act.
Type:	string, list of keywords separated by a comma
Examples:	EU employment policy, working conditions, fight against unemployment
Note:	EuroVoc is a dictionary of standardised keywords used to describe legal documents. These tags are organised in 21 domains and 127 sub-domains. Please refer to the complete list of keywords at the Eur-Lex website.

Variable name:	`act_directory`
Description:	Directory of the adopted act.
Type:	string
Example:	Social policy → Employment and unemployment → Protection of workers
Note:	Please refer to the full tree at the Eur-Lex website.

Variable name:	`act_date`
Description:	Date of publication the adopted act.
Type:	numeric, date format %td
Example:	22nov2024

Variable name:	`act_author`
Description:	Author of the document.
Type:	string
Example:	European Parliament, European Commission

Variable name:	`act_effect`
Description:	Date of effect of the adopted act.
Type:	numeric, date format %td
Example:	22nov2024

2.2 Procedures

File name: `procedures.dta`
Description: This dataset contains all EU inter-institutional procedures regarding social policy, education and training.
Unit of observation: Procedure.

Variable name:	`procedure_number`
Description:	Legislative procedure number. Linking variable
Type:	string
Example:	2023/0179/NLE
Note:	Procedure numbers have three parts: the year, the procedure number (digits and/or letters) and abbreviation of the procedure type. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`procedure_date`
Description:	Date of procedure
Type:	numeric, date format %td
Example:	22nov2024

Variable name:	`procedure_form`
Description:	Form of the procedure.
Type:	numeric (labeled)
Label list:	1 Communication
	2 Draft decision
	3 Draft recommendation
	4 Draft regulation
	5 Legislative resolution
	6 Other EP procedures
	7 Proposal for a decision
	8 Proposal for a decision without addressee
	9 Proposal for a directive
	10 Proposal for a recommendation
	11 Proposal for a regulation
	12 Proposal for an act
	13 Proposal for an implementing decision
	14 Resolution

Variable name:	`procedure_type`
Description:	Type of procedure.
Type:	numeric (labeled)
Label list	1 Agreement
	2 Budgetary procedure
	3 Cooperation procedure
	4 Legislative initiative procedure
	5 Non-legislative procedure
	6 Ordinary legislative procedure

	7 Own-initiative procedure 8 Resolutions on topical subjects 9 Social protocol 10 Special legislative procedure (EP consent required) 11 Special legislative procedure (EP consulted)

Variable name:	`procedure_status`
Description:	Procedure status
Type:	numeric (labeled)
Label list:	0 Ongoing 1 Completed 2 Withdrawn

Variable name:	`act?`
Description:	CELEX number of the legal act originated from the procedure. Linking variables 🔗
Type:	string
Example:	32023H01344
Note:	Some procedures have originated several legal acts. This variable appears several times in the dataset, numbered as act1, act2, act3, act4. This information was arranged like this to comply with the structure of the dataset, where one observation corresponds to a procedure. To rearrange the data in Stata, use the command `reshape wide`.

2.3 European Parliament resolutions

File name: `EP_resolutions.dta`
Description: This dataset contains all European Parliament resolutions (preparatory documents) that concern either adopted acts or procedures in the areas of social policy, education and training.
Unit of observation: Resolution.

Variable name:	`resolution_name`
Description:	Name of the EP resolution.
Type:	string
Example:	European Parliament resolution of 1 June 2023 on strengthening social dialogue.

Variable name:	`resolution_number`
Description:	CELEX number of the EP resolution.
Type:	string
Example:	52023IP0218
Note:	CELEX numbers are unique identifiers used for European Union legal documents. They have four parts: the legal sector, the year of publication, the category of legal act (e.g., directive, regulation), and a unique document number. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`procedure_number`
Description:	Legislative procedure number. Linking variable
Type:	string
Example:	2023/2536/RSP
Note:	Procedure numbers have three parts: the year, the procedure number (digits and/or letters) and abbreviation of the procedure type. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`resolution_form`
Description:	Form of EP resolution.
Type:	numeric (labeled)
Label list:	1 Acceptance of common position 2 Approved joint text 3 Legislative resolution 4 Opinion not proposing amendment 5 Opinion proposing amendment 6 Own-initiative report 7 Own-initiative resolution 8 Recommendation 9 Resolution

Variable name:	`resolution_eurovoc`
Description:	EUROVOC keywords of the EP resolution.
Type:	string, list of keywords separated by a comma
Examples:	EU employment policy, working conditions, fight against unemployment
Note:	EuroVoc is a dictionary of standardised keywords used to describe legal documents. These tags are organised in 21 domains and 127 sub-domains. Please refer to the complete list of keywords at the Eur-Lex website.

Variable name:	`term`
Description:	Parliamentary term.
Type:	numeric (labeled)
Label list:	5 1999-2004
	6 2004-2009
	7 2009-2014
	8 2014-2019
	9 2019-2024
	10 2024-2029

Variable name:	`resolution_date`
Description:	Date of the document.
Type:	numeric, date format %td
Example:	22nov2024

Variable name:	`vote_date`
Description:	Date of vote.
Type:	numeric, date format %td
Example:	22nov2024

2.4 Transpositions of adopted acts to national law

File name: `transpositions.dta`
Description: This dataset contains all transpositions of EU adopted acts regarding social policy, education and training to national policy.
Data structure: Adopted act > Transposition of acts

Variable name:	`country`
Description:	Country
Type:	numberic (labeled)
Label list:	1 Austria
	2 Belgium
	3 Bulgaria
	4 Croatia
	5 Cyprus
	6 Czechia
	7 Denmark
	8 Estonia
	9 Finland
	10 France
	11 Germany

12 Greece
13 Hungary
14 Ireland
15 Italy
16 Latvia
17 Lithuania
18 Luxembourg
19 Malta
20 Netherlands
21 Poland
22 Portugal
23 Romenia
24 Slovakia
25 Slovenia
26 Spain
27 Sweden
28 United Kingdom

Variable name:	`geo`
Description:	Country: alpha-2 country code
Type:	string
Example:	NL

Variable name:	`act_number`
Description:	CELEX number of the adopted act. Linking variable
Type:	string
Example:	32023H01344
Note:	CELEX numbers are unique identifiers used for European Union legal documents. They have four parts: the legal sector, the year of publication, the category of legal act (e.g., directive, regulation), and a unique document number. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`act_date`
Description:	Date of the legal act.
Type:	numeric, date format %td
Example:	22nov2024

Variable name:	`transposition_date`
Description:	Date of the document transposing the EU legal act to domestic law.
Type:	numeric, date format %td
Example:	22nov2024

Variable name:	`transposition_name`
Description:	Name of the document transpositing the EU legal act to domestic law.
Type:	string
Example:	Parent's Leave and Benefit Act 2019 (Extension of Periods of Leave) Order 2024 (Ireland)

2.5 Voting records

File name: `voting_records.dta`
Description: This dataset includes voting records for all the European Parliament resolutions pertraining adopted acts related to social policy, education and training. These records refer to roll-call votes only. In case there are multiple voting records for a single EP resolution (for instance, because several amendments are voted separatedely), the dataset includes instances of *single vote*.
Data structure: EP resolution > Vote

Variable name:	`resolution_number`
Description:	CELEX number of the EP resolution.
Type:	string
Example:	52023IP0218
Note:	CELEX numbers are unique identifiers used for European Union legal documents. They have four parts: the legal sector, the year of publication, the category of legal act (e.g., directive, regulation), and a unique document number. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website.

Variable name:	`mep`
Description:	Name of the Member of Parliament. Linking variable
Type:	string
Example:	Fiona HALL

Variable name:	`name`
Description:	Original name in the voting records.
Type:	string
Example:	Hall

Variable name:	`party_acronym`
Description:	Party Family (original acronym)
Type:	string
Example:	ALDE

Variable name:	`party_family`
Description:	Party Family. Standardised party family across legislative terms.
Type:	numeric (labeled)
Label list:	1 Left 2 Greens 3 Social Democrats 4 Liberals 5 Christian Democrats 6 Conservatives 7 Far right 8 Non-attached members

Variable name:	`vote`
Description:	Vote
Type:	numeric (labeled)
Label list:	0 Against 1 Abstention 2 For

Variable name:	`term`
Description:	Parliamentary term.
Type:	numeric (labeled)
Label list:	7 2009-2014 8 2014-2019 9 2019-2024

2.6 Explanations of votes

File name: `explanations.dta`

Description: This dataset contains all recorded explanations of the votes of MEPs regarding the European Parliament resolutions included in the voting records.
Data structure: EP resolution > MEP > Explanation of vote

Variable name:	`mep`
Description:	Name of the Member of Parliament. Linking variable 🔗
Type:	string
Example:	Fiona HALL

Variable name:	`name`
Description:	Original name on the European Parliament website.
Type:	string
Example:	Fiona Hall

Variable name:	`party_family`
Description:	Party Family. Standardised party family across legislative terms.
Type:	numeric (labeled)
Label list:	1 Left 2 Greens 3 Social Democrats 4 Liberals 5 Christian Democrats 6 Conservatives 7 Far right 8 Non-attached members

Variable name:	`resolution_number`
Description:	CELEX number of the EP resolution.
Type:	string
Example:	52023IP0218
Note:	CELEX numbers are unique identifiers used for European Union legal documents. They have four parts: the legal sector, the year of publication, the category of legal act (e.g., directive, regulation), and a unique document number. Please refer to the detailed explanation of how these codes are constructed at the Eur-Lex website 🔗.

Variable name:	`explanation`
Description:	Explanation of vote
Type:	string
Example:	I voted in support of this motion as it supports the reintegration in the labour market of workers who have been made redundant and works to decrease unemployment within the affected area and social grouping.

Variable name:	`language`
Description:	Identified language of the explanation of vote.
Type:	numeric (labeled)
Notes:	The European Parliament website has no information on the language of explanations of votes. This variable was constructed using language detection module in Python called `langdetect`.
Label list:	1 Bulgarian 2 Croatian 3 Czech 4 Danish 5 Dutch 6 English 7 Estonian 8 Finnish 9 French 10 German 11 Greek 12 Hungarian 13 Irish 14 Italian 15 Latvian 16 Lithuanian 17 Maltese 18 Polish 19 Portuguese 20 Romanian 21 Slovakian 22 Slovenian 23 Spanish 24 Swedish

Variable name:	`term`
Description:	Parliamentary term.
Type:	numeric (labeled)
Label list:	7 2009-2014
	8 2014-2019
	9 2019-2024

2.7 Information about members of Parliament

File name: `MEPs.dta`
Description: This dataset includes information about Members of Parliament for each legislative term.
Data structure: Legislative term > MEP

Variable name:	`mep`
Description:	Name of the Member of Parliament. Linking variable 🔗
Type:	string
Example:	Fiona HALL

Variable name:	`term`
Description:	Parliamentary term.
Type:	numeric (labeled)
Label list:	7 2009-2014
	8 2014-2019
	9 2019-2024

Variable name:	`party_family`
Description:	Party Family. Standardised party family across legislative terms.
Type:	numeric (labeled)
Notes:	Observations correspond to the party family of MEPs at the time of the previous European election. These observations may be different than those in the voting records and explanations dataset, which refer to the party family at the time the vote was casted or the speech was given.
Label list:	1 Left
	2 Greens
	3 Social Democrats
	4 Liberals
	5 Christian Democrats
	6 Conservatives
	7 Far right

8 Non-attached members

Variable name: `dob`
Description: Date of birth
Type: numeric, date format %td
Example: 18apr1958

Variable name: `yob`
Description: Year of birth
Type: numeric
Example: 1958

Variable name: `partyfamily_original`
Description: Party family (original coding on the website for the MEP)
Type: string
Example: Left, The Left group in the European Parliament

Variable name: `country`
Description: Country
Type: numeric (labeled)
Label list: 1 Austria
2 Belgium
3 Bulgaria
4 Croatia
5 Cyprus
6 Czechia
7 Denmark
8 Estonia
9 Finland
10 France
11 Germany
12 Greece
13 Hungary
14 Ireland
15 Italy
16 Latvia
17 Lithuania
18 Luxembourg
19 Malta
20 Netherlands

21 Poland
22 Portugal
23 Romania
24 Slovakia
25 Slovenia
26 Spain
27 Sweden
28 United Kingdom

Variable name:	`geo`
Description:	Country: alpha-2 country code
Type:	string
Example:	NL

Variable name:	`legislature?`
Description:	Dummy variable identifying whether the politician in question was an MEP in the first/second/third.../nineth legislature.
Type:	numeric
Note:	This variable appears several times in the dataset, numbered as legislature1, legislature2, legislature3, ..., legislature9.
	Legislature 1: 1979–1984
	Legislature 2: 1984–1989
	Legislature 3: 1989–1994
	Legislature 4: 1994–1999
	Legislature 5: 1999–2004
	Legislature 6: 2004–2009
	Legislature 7: 2009-2014
	Legislature 8: 2014–2019
	Legislature 9: 2019–2024

Variable name:	`committee_*`
Description:	Role in parliament committee.
Type:	numeric (labeled)
Label list:	0 Not a member
	1 Member
	2 Vice-chair
	3 Chair
	This variable appears several times in the dataset for each committee in the EP.
`committee_agriculture`	Committee on Agriculture and Rural Development
`committee_budget`	Committee on Budgets

committee_jha	Committee on Civil Liberties, Justice and Home Affairs
committee_constitution	Committee on Constitutional Affairs
committee_education	Committee on Culture and Education
committee_development	Committee on Development
committee_economy	Committee on Economic and Monetary Affairs
committee_social	Committee on Employment and Social Affairs
committee_fisheries	Committee on Fisheries
committee_foreign	Committee on Foreign Affairs
committee_industry	Committee on Industry, Research and Energy
committee_trade	Committee on International Trade
committee_legal	Committee on Legal Affairs
committee_petitions	Committee on Petitions
committee_regions	Committee on Regional Development
committee_transport	Committee on Transport and Tourism
committee_gender	Committee on Women's Rights and Gender Equality
committee_environment	Committee on the Environment, Public Health and Food
committee_consumer	Committee on the Internal Market and Consumer Protection

Variable name:	national_party
Description:	National party
Type:	string
Example:	Democratic Party

3 Linking datasets

3.1 Direct links

Adopted acts → procedures:

```
merge m:1 procedure_number using "procedures.dta", keep(1 3)
gsort -document_date
```

EP resolutions → procedures:

```
merge m:1 procedure_number using "procedures.dta", keep(1 3)
gsort -resolution_date
```

Transpositions of vote → adopted acts:

```
merge m:1 document_number using adopted_acts, keep(1 3)
gsort -document_date -transposition_date
```

Voting records → MEPs:

```
merge m:1 mep term using MEPs.dta, keep(1 3)
gsort -resolution_number
```

Explanations of vote → MEPs:

```
merge m:1 mep term using MEPs.dta, keep(1 3)
gsort -resolution_number
```

Explanations of vote → voting records:

```
merge m:1 mep resolution_number using voting_records.dta, keep(1 3)
gsort -resolution_number
```

3.2 Other combinations

These linking variables can be used to achieve additional combinations of data merging. The procedure is straightforward: begin with the dataset containing the most disaggregated level of observations and progressively merge upwards. Below is a reference list, ordered from the lowest to the highest level of data aggregation:

EP data	Document data
1. Explanations of vote	1. Transposition of votes
2. Vote	2. Adopted acts
3. EP Resolutions	3. Procedures
4. Procedures	

Any combination within the same column can be merged through consecutive merge operations. For example, to combine (1) explanations of votes with (2) voting records and (3) EP resolutions, follow this procedure:

```
use explanations

merge m:1 mep resolution_number using voting_records, keep(1 3)
drop _merge

merge m:1 resolution_number using EP_resolutions, keep(1 3)
gsort -resolution_number
```

Merging operations across different columns require a different approach. This is because a single adopted act may correspond to multiple EP resolutions as preparatory documents. Similarly, an EP resolution within a specific policy process may result in multiple adopted acts. Thus, a simple `m:1` merge cannot be used to link these types of data. Instead, the best practice is to use the `joinby` command, as illustrated below:

```
use EP_resolutions

joinby procedure_number using adopted_acts, unmatched(master)
gsort -resolution_number
```

References

Eur-Lex. (2024). *Adopted acts, procedures, and European Parliament resolutions*. Retrieved 16/09/2024, from https://eur-lex.europa.eu/homepage.html

Golub, J. (2024). EUPROPS: A new dataset on policymaking in the European Union from 1958 to 2021. *European Union Politics*, *25*(1), 197–217.

Legislative Observatory of the European Parliament. (2024). *Voting records and explanations of vote*. Retrieved 16/09/2024, from https://oeil.secure.europarl.europa.eu/oeil/en

Parltrack. (2024). *Voting records, legislative procedures*. Retrieved 20/11/2024, from https://parltrack.org