



Deliverable D4.1 – Two datasets on objective descriptive and substantive representation with a particular focus on gender equality, climate change and migration

PROJECT	
Project number:	101094190
Project acronym:	ActEU
Project name:	Towards a new era of representative democracy - Activating European citizens' trust in times of crises and polarization
Call:	HORIZON-CL2-2022-DEMOCRACY-01
Topic:	HORIZON-CL2-2022-DEMOCRACY-01-08
Type of action:	HORIZON-RIA
Responsible service:	REA
Project starting date:	01 / 03 / 2023
Project duration:	36 months

DELIVERABLE	
Deliverable No	14
Deliverable related No	4.1
Work package No	WP4
Description	Two datasets on objective descriptive and substantive representation with a particular focus on gender equality, climate change and migration
Lead beneficiary	13. Agencia Estatal Consejo Superior de Investigaciones Científicas (CSIC)
Type	DATA
Dissemination level	PU - Public
Due to date	M19 – 30.09.2024
Main author(s): ¹	Laura Morales, Carles Pamies, Nayla Escribano (CSIC)
Contributor(s): ²	Alexandros Gkotinakos (AUTH), Matilde Ceron (PLUS), Ermela Gianna (PLUS), Eftychia Teperoglou (AUTH), Zoe Lefkofridi (PLUS), Nadine Zwiener-Collins (PLUS)
Approved by	Project Coordinator Prof. Dr. Michael Kaeding (UDE)

¹ The authors have written the report, designed data collection and coordinated data collection efforts by all teams contributing to the work described in this deliverable.

² The contributors have contributed with feedback to data collection design proposals by the leading team, quality checks of data collection instruments (such as the translation of keyword dictionaries) and contributed data that is described with their respective data citations in this document.



Two Datasets on Objective Descriptive and Substantive Representation

Executive Summary and Key Recommendations

Summary of context

The ActEU project examines the decline of trust and of democratic political legitimacy in Europe through a conceptual triangle that proposes that fully understanding the magnitude and nature of this decline requires examining closely interconnected factors linked to citizens' political orientations, citizens' political participation and how representative institutions and actors operate. The role of WP4 in the context of the ActEU project is to assess how political representation dynamics shape the manifestations of the crisis of trust and legitimacy.

In order to do this, several tasks and deliverables focus on descriptive and substantive political representation as dimensions that reflect how the political system and elected representatives “objectively” represent citizens. The goal of this deliverable is to provide datasets on descriptive and substantive political representation that allow gauging (i) how do the current political and social developments affect the representation of citizens’ preferences, and (ii) for which social groups of society do the current political and social developments have tangible implications in terms of the representation of their preferences. Moreover, following the dedicated focus of the ActEU project on three policy and public areas of debate around climate change, gender equality and immigration, the datasets allow to capture both the political and social groups particularly interested in these policy fields as well as the representation outputs of parliaments and parties in relation to these three fields.

Added value

As the report will outline, although social scientists are increasingly paying attention to issues around democratic political representation, there is limited comparative data available on descriptive and substantive representation across Europe that is also rich in information about both dimensions of political representation. This report adds value to the existing state of knowledge with three contributions. First, it takes stock of the data that exists and their limitations. Second, it provides two novel datasets for a selection of the countries on which the ActEU project focuses, for a comprehensive set of indicators on both descriptive and substantive representation aggregated at the country/legislature/party level with special attention to climate change, gender equality and immigration.

Summary of Deliverable D4.1

This report is structured in five main sections. The first section provides an overall introduction to conceptual and theoretical aspects relevant to the study of democratic political representation. Section two takes stock of the existing research and datasets on descriptive and substantive representation, pointing out their limitations. The third section describes the methodology



followed for the creation of the dataset on descriptive representation, whereas section four outlines the methodology followed for the creation of the dataset on substantive representation. The final section offers some policy-relevant recommendations that we also briefly summarise here.

Key conclusions and policy recommendations

The main goal of this deliverable and the tasks associated to it was not to produce findings and recommendations, but data. However, in the process of reviewing the existing data we are able to make some policy-relevant recommendations. Despite the key importance of both descriptive and substantive representation for our understanding of political dynamics in contemporary democracies, the existing data is too fragmentary and of uneven quality. We have detected numerous limitations of the existing datasets on legislators' biographies regarding the coverage of a variety of socio-demographic, socio-economic and socio-political groups/categories. Moreover, the absence of standardised approaches regarding how national parliaments (and other sub-national elected assemblies) provide open data on (i) the elected representatives and their characteristics/attributes, (ii) their parliamentary roles, positions and activities, and (iii) the numerous speeches, documents and agreements produced in these parliaments make EU-wide comparisons difficult.

The main recommendations we make are: (1) DG Research could foster a set of task forces or initiatives that would contribute to more standardised approaches to Open Data from legislative assemblies, perhaps with the collaboration of Eurostat, where the expertise of political scientists, data scientists and computational linguists can be combined to provide with harmonised guidelines or good practice recommendations as well as ear-marked funding to improve adaptation to new standards; (2) Existing infrastructures such as CLARIN, CESSDA and LIBER could support the technical work required; (3) Collaborations with the Inter-Parliamentary Union could be highly productive to scale up these initiatives beyond the EU.



Content

Figures/Tables	4
1 Introduction	5
2 Studying descriptive and substantive representation: Taking stock of existing data	7
2.1 Existing research and datasets on descriptive representation in Europe	7
2.1.1 List of publications and datasets on descriptive representation in Europe mapped for the ActEU project	8
2.1.2 Summary of key attributes of the main comparative datasets on descriptive representation in Europe mapped for the ActEU project	10
2.2 Existing research and datasets on substantive representation in Europe	15
2.2.1 List of publications and datasets on substantive representation in Europe mapped for the ActEU project	15
2.2.2 Summary of key attributes of the main comparative datasets on substantive representation in Europe mapped for the ActEU project	19
3 The ActEU datasets on descriptive political representation	22
3.1 Data collected, methodology and variables	23
3.2 Descriptive statistics of the datasets collected	26
4 The ActEU dataset on substantive political representation	27
4.1 Data collected, methodology and variables	27
4.2 Descriptive statistics of the datasets collected	31
5 Conclusions and policy-relevant recommendations	32
Bibliography/List of References	33
Appendix - Links to codebooks, templates and other technical documentation of the datasets	35
A.1 Codebooks and templates used for the descriptive representation datasets	35
A.2 Codebooks and templates used for the substantive representation datasets	35
A.3 List of keywords in English used for the substantive representation datasets	36



Figures/Tables

Table 1 Summary of the main comparative datasets on descriptive representation found through the systematic search. (Carles Pamies & Laura Morales, CSIC)	11
Table 2 Summary of the variables included in the main comparative datasets on descriptive representation useful for ActEU found through the systematic search. (Carles Pamies & Laura Morales, CSIC)	13
Table 3 Summary of datasets on substantive representation found through the systematic search. (Carles Pamies & Laura Morales, CSIC)	20
Table 4 Summary of the variables included in the main comparative datasets on substantive representation useful for ActEU found through the systematic search. (Carles Pamies & Laura Morales, CSIC)	21
Table 5 Description of the data collected (Laura Morales and Carles Pamies, CSIC)	23
Table 6 Description of the variables included in the descriptive representation dataset (Laura Morales and Carles Pamies, CSIC)	24
Table 7 Descriptive statistics of the key variables included in the descriptive representation datasets (Laura Morales and Carles Pamies, CSIC)	26
Table 8 Description of the data collected on substantive representation (Laura Morales and Nayla Escibano, CSIC)	28
Table 9 Description of the variables included in the substantive representation datasets (Laura Morales and Nayla Escibano, CSIC)	29
Table 10 Descriptive statistics of the variables included in the substantive representation datasets (Laura Morales and Nayla Escibano, CSIC)	31
Table 11 Description of the variables included in the descriptive representation dataset (Laura Morales and Nayla Escibano, CSIC)	36



1 Introduction

Representative democracies across Europe are experiencing a period of crisis. A considerable proportion of European citizens show a lack of trust in the core institutions and actors that make representative democracy work at various levels of government (cf. e.g., Van der Meer & Van Erkel 2024; Valgarðsson et al. 2024; Carstens 2023; Norris 2023; Torcal & Christmann 2021). Although the downward trend in political trust is not universal across countries (see Christensen et al. 2024), it is undeniable that in many EU Member States political trust in all actors and institutions is in decline. This crisis in political trust, unavoidably, is associated to feelings of bad or misrepresentation: when citizens do not trust politicians, parties, the legislative or the executive – to name but a few – they tend not to feel represented by these actors and institutions, and the other way around (Bedock & Panel 2017; Jacobs, Close & Pilet 2024; Palacios 2016). Other reports and outputs of the ActEU project focus on these connections between political trust and feelings of representation. Here we focus on cross-national variations and dynamics of political representation as an object of analysis.

Indeed, the ActEU project focuses on several facets of the process of democratic political representation: on the one hand, a more 'objective' facet consisting of (i) the descriptive representation achieved by European parliaments and the parties represented therein by virtue of the selection of legislators with a range of socio-economic and demographic attributes, and (ii) the substantive representation achieved by parliaments and parties through their parliamentary activities; and, on the other hand, a more 'subjective' facet that pertains to the perceptions that citizens express of how well or badly parliaments and parties represent them and their demands. In this report, we present several datasets that allow us to assess the first more 'objective' or 'objectifiable' facet.

In essence, we focus on what Hanna Pitkin (1967) defined as descriptive and substantive representation. Elected representatives can be said to 'descriptively' represent citizens when their shared attributes with the latter allow them to 'stand for' on their behalf. By mirroring (some of) their characteristics (e.g., sex, age, ethnicity, class, various identities, etc.) to a greater or less extent, the elected representatives and the aggregation of all of them (i.e., elected assemblies) will be providing a more or less accurate reflection of the citizenry they are expected to represent. In this sense, some parliaments and political parties will better represent descriptively the societies that elected them, others will do so more imperfectly. Of course, how well any given representative body descriptively represents society as a whole can be judged according to a wide range of characteristics and attributes. In this report, we limit this assessment to a few attributes: apparent sex, age, minority background, gender identity and sexual orientation.

Substantive representation, by contrast, is said to happen when elected representatives act in a manner that they 'act for' those who elected them. This form of representation requires representatives to reasonably voice the concerns and demands that those who elected them care for. Assessing substantive representation is not always as easy as it sounds, as it requires first establishing what citizens want and prefer (and citizens' demands and preferences are heterogeneous, by and large) and then examining how much these preferences and demands are present in parliamentary activities (and parliamentary activities are also heterogeneous and leading to multiple types of outputs). In this report, we limit this assessment to the three main areas of policy-making on which the ActEU project focuses particularly (climate change, gender



equality and LGBTIQ+ rights, and immigration) and add youth-related issues and minority-related issues to match the social categories examined in the descriptive representation datasets.

The main goal of this deliverable is, thus, to take stock of existing data on both descriptive and substantive representation, and to describe the contents and methodology employed to generate two novel aggregate datasets on descriptive and substantive representation that can be used within the ActEU project to link (in subsequent reports and publications) citizens' perceptions and feelings (as measured, for example, by the main ActEU survey) with these more 'objectifiable' outputs of democratic representation regarding the descriptive and substantive facets of representation.

Nevertheless, our work on pursuing this main goal also allows us to provide some policy-relevant recommendations, especially regarding future Open Data availability and production by European parliaments and joint research capacity in the study of parliamentary data and corpora.

The report is structured as follows. The next section takes stock of the existing data on descriptive and substantive representation across Europe, focusing particularly on cross-national datasets and studies. Section three outlines the methodology followed to generate the datasets on descriptive representation and explains their exact contents. Section four follows in the same manner presenting the methodology and contents of the datasets on substantive representation. The final section concludes with a summary of the work undertaken and submitted with the datasets, its advantages and limitations, and with policy-relevant recommendations.



2 Studying descriptive and substantive representation: Taking stock of existing data

Before generating novel data, the Sciences Po and CSIC teams³ undertook an extensive review of the state of the art on data collection regarding both descriptive and substantive representation. While research on these topics has been mainly undertaken by political scientists and political sociologists, computational linguistics interested in parliamentary corpora have also contributed notable datasets to this subfield of research on legislative politics.

The scoping exercise was organised as follows. First, we searched for and considered any and all datasets on either descriptive representation (biographical information on legislators) and substantive representation (activities of the legislators) for three levels of democratic representation since the early 2000s: the European Parliament, national parliaments and subnational legislative assemblies (states in federal systems and regions in quasi-federal or decentralised systems). To this end, we undertook a systematic search on Google Scholar, scientific publications databases as well as data repositories (such as Harvard Dataverse and the archives in CESSDA) using keywords relating to “parliament”, “descriptive representation”, “substantive representation” and “legislators”. We also used snowball techniques following citations in work that referred to other publications or datasets.

Second, we inspected each of those articles or dataset reports to examine what they contained in terms of the variables included (attributes or activities of the legislators) and recorded their contents. In particular, we assessed whether the datasets contained information at the individual-level of the legislators or they only provided aggregates (whether for the parliament/assembly as a whole or aggregated by political party or other similar political unit). This allowed us to assess the utility of the datasets for our purposes, as individual-level data provided us with greater flexibility for re-analysis and restructuring.

Finally, we assessed which of the datasets were available as Open Data for reuse and which remained private data that was not shared or sufficiently documented for reuse.

Section 2.1 presents the results of this stock-taking of existing data for descriptive representation, while section 2.2 presents the same exercise for substantive representation.

2.1 Existing research and datasets on descriptive representation in Europe

Through the aforementioned systematic search process, we uncovered a wide range of publications and/or datasets that include biographical information on legislators that is potentially of use for the ActEU project. We first present the list of publications and datasets and then summarise their main attributes for a selection of those available for public use.

³ Work on this deliverable and the associated tasks started when the team now based at CSIC was based in Sciences Po, hence the reference to both consortium partners throughout the document.



2.1.1 List of publications and datasets on descriptive representation in Europe mapped for the ActEU project

- Auer, Daniel, Portmann, Lea, & Tichelbaecker, Tichelbaecker (2023). Electoral Discrimination, Party Rationale, and the Underrepresentation of Immigrant-Origin Politicians. *American Journal of Political Science* (Online First). <https://doi.org/10.1111/ajps.12817>
- Azavea (2018). *Cicero*. Available from <https://www.cicerodata.com>
- Bailer, Stefanie, Manow, Philip, & Hug, Simon (2019). *Parliamentary Careers in Comparison: Research Project*. Basel, Bremen, Geneva: Retrieved from <http://parliamentarycareersincomparison.org/>
- Best, Heinrich & Cotta, Maurizio (2000). *Parliamentary representatives in Europe 1848-2000*. Oxford: Oxford University Press.
- Best, Heinrich & Edinger, Michael (2005). Converging representative elites in Europe? An introduction to the EurElite project. *Czech Sociological Review*, 41(3), 499-510. <https://www.jstor.org/stable/41132162>
- Costa, Monica(2016). Gender-Focused Institutions in International Parliamentary Bodies: The Case of the Women's Caucus of the Parliamentary Assembly of the Portuguese-Speaking Countries. *Parliamentary Affairs*, 69(4), 748-762. <https://doi.org/10.1093/pa/gsw002>
- Elsässer, Lea & Schäffer, Armin (2022). *Unequal Democracies: Who does (not) have a seat in parliament?* Friedrich Ebert Stiftung. <https://library.fes.de/pdf-files/bueros/wien/19349-20220926.pdf>
- Gerring, John, Onel, Erzen, Morrison, Kevin, and Pemstein, Daniel (2019). Who Rules the World? A Portrait of the Global Leadership Class, *Perspectives on Politics*, 17(4), 1079-1097. <https://doi.org/10.1017/S1537592719000744>. Dataset available here: <http://glp.la.utexas.edu/>
- Göbel, Sascha & Munzert, Simon (2022). The comparative legislators database. *British Journal of Political Science*, 52(3), 1398-1408. <https://doi.org/10.1017/S0007123420000897>. Also available here: <https://complegdatabase.com/>
- Gorecki, Maciej A., & Pierzgalski, Michał. (2022). Legislated candidate quotas and women's descriptive representation in preferential voting systems. *European Journal of Political Research*, 61(1), 154-174. <https://ejpr.onlinelibrary.wiley.com/doi/full/10.1111/1475-6765.12443?campaign=woletoc>
- Gouglas, Athanassios, Bart Maddens and Marleen Brans (2018). "Determinants of legislative turnover in Western Europe, 1945-2015." *European Journal of Political Research*, 57(3):637-661. <https://doi.org/10.1111/1475-6765.12246>



- Inter-Parliamentary Union (2024). *PARLINE Database on National Parliaments* <https://data.ipu.org/>
- Lupu, Noam, & Warner, Zach (2022). Affluence and congruence: unequal representation around the world. *The Journal of Politics*, 84(1), 276-290. <https://doi.org/10.1086/714930>
- Lupu, Noam & Warner, Zach (2022). Why are the affluent better represented around the world? *European Journal of Political Research*, 61(1), 67-85. <https://doi.org/10.1111/1475-6765.12440>
- Matland, Richard E. and Donley T. Studlar. 2004. Determinants of legislative turnover. A cross-national analysis." *British Journal of Political Science*, 34(1):87– 108.
- Michon, Sébastien, & Wiest, Eric (2021). A database about the Members of European Parliament: Contributions and limitations of automated data collection in the study of European political elites. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 152(1), 125-141. <https://doi.org/10.1177/07591063211040235>
- Morales, Laura; Vintila, Daniela; Geese, Lucas; Mügge, Liza; van der Pas, Daphne; van de Wardt, Marc (2017). *Codebook and data collection guidelines of work package 1 on descriptive political representation in national parliaments of the project Pathways to Power*. <https://doi.org/10.7910/DVN/MRKJJF>, Harvard Dataverse, V1
- Morales, Laura; Vintila, Daniela; Geese, Lucas; Mügge, Liza; van der Pas, Daphne; van de Wardt, Marc (2017). *Codebook and data collection guidelines of work package 1 on descriptive political representation in regional parliaments of the project Pathways to Power*, <https://doi.org/10.7910/DVN/KZMOAM>. Harvard Dataverse, V1
- MySociety (2018). *EveryPolitician*. Available from <https://everypolitician.org>
- Nyrup, Jacob, & Bramwell, Stuart (2020). Who governs? A new global dataset on members of cabinets. *American Political Science Review*, 114(4), 1366-1374. <https://doi.org/10.1017/S0003055420000490>. Dataset and documentation available here: <https://politicscentre.nuffield.ox.ac.uk/whogov-dataset/download-dataset/>
- Reynolds, Andrew (2013). Representation and rights: The impact of LGBT legislators in comparative perspective. *American Political Science Review*, 107(2), 259-274. <https://www.jstor.org/stable/43654014>
- Salvati, Eugenio (2022). *Dataset on Members of the European Parliament (1979-2019)* [Dataset, version 5.1]. Harvard Dataverse. <https://doi.org/10.7910/DVN/V2FJEF>
- Rayside, David. 2018 [1998]. *On the Fringe: Gays & Lesbians in Politics*. Ithaca, NY: Cornell University Press
- Transparency International EU (2023). *EU Integrity Watch*. <https://data.integritywatch.eu/> [Note: this also includes the information on MEPs income, <https://www.integritywatch.eu/mepincomes>]



- Vanlangenakker, Ine, Wauters, Bram, & Maddens, Bart (2013). Pushed toward the exit? How female MPs leave parliament. *Politics & Gender*, 9(1), 61-75.
<https://doi.org/10.1017/S1743923X12000694>
- Vincent-Mory, Claire & Morales, Laura (2024). *Codebook and Data Collection Guidelines of the Project InclusiveParl on the Descriptive Representation of LGBTI+ Individuals*. <https://doi.org/10.7910/DVN/HBKBM0>. Harvard Dataverse, V1.
- Wagner, Claudia (2017). *Politicians on Wikipedia and DBpedia*. GESIS, Cologne. Data File Version 1.0.0, <https://doi.org/10.7802/1515>

2.1.2 Summary of key attributes of the main comparative datasets on descriptive representation in Europe mapped for the ActEU project

After this mapping exercise, we have selected the main datasets that provided comparative biographical data on legislators across several countries in Europe and which could be of use for the ActEU project as sources.

Table 1 provides this summarised outlook with several key pieces of information: the period covered, whether the dataset provides individual-level data for each legislator in the given assembly, whether the dataset is available to ActEU researchers either through existing public data repositories or because of their participation in the respective projects, and which of the ActEU ten countries of focus are covered by the given dataset. Additionally, Table 2 shows the range of variables included in each of the main studies considered during the mapping and scoping stage.

As can be seen, both the country coverage, time coverage and variable coverage are very uneven, resulting in the impossibility of resorting to available datasets for readily usable data for all ActEU countries.



Table 1 Summary of the main comparative datasets on descriptive representation found through the systematic search. (Carles Pamies & Laura Morales, CSIC)

Database	Period covered	Individual level data	Access to data?	CZ	ES	FR	IT	PL	DE	FI	DK	AT	GR
<u>European</u>													
Integrity Watch EU	2014-	Yes	Using online platform	2014-	2014-	2014-	2014-	2014-	2014-	2014-	2014-	2014-	2014-
Lupu and Warner (2022)	1967-2015	Yes	Only aggregate	2004-2015	1996 + 1999-2015	1967 + 1996-2015	1996 + 1999-2015	2004-2015	1996 + 1999-2015	1996 + 1999-2015	1996 + 1999-2015	1996 + 1999-2015	1987-2015
Members of the European Parliament	1979 - 2019	Yes	Yes	2004-2019	1986-2019	1979-2019	1979-2019	2004-2019	1979-2019	1995-201	1979-2019	1995-2019	1981-2019
<u>National</u>													
Comparative legislators database	1789 - 2022	Yes	Yes	1992-2021	1979-2019	1958-2022	1948-2022		1949-2021	.	.	1920-2019	.
EurElite	1814 - 2004	Unknown, probably not	No	1992-2004	1868-1917 + 1977-1979 + 1996-2004	1871-1940 1948-2004	1861-1928 + 1946-2004	1989-2004	1849-2004	1907-2003	1849-2004	1918-1933 + 1945-1995	.
Every Politician	1920 - 2018	Yes	Yes	1992-2018	2011-2016	2002-2018	2013-2018	1991-2018	1949-2021	1970-2018	1990-2019	2013-2017	1974-2018
Global Leadership Project	2013 - now	Yes	Yes	2013-	2013-	2013-	2013-	2013-	2013-	2013-	2013-	2013-	2013-
Gouglas et al. (2018)	1945-2015	Unknown	Only aggregate	.	.	1945-2015	1945-2015	.	1945-2015	1945-2015	1945-2015	1945-2015	1945-2015
InclusiveP	2008 -	Yes	Yes	.	2008-2023	2012-2017	2013-2022



arl / Repchance	2023												
Lupu and Warner (2022)	1967-2015	Yes	Only aggregate	2004-2015	1996 + 1999-2015	1967 + 1996-2015	1996 + 1999-2015	2004-2015	1996 + 1999-2015	1996 + 1999-2015	1996 + 1999-2015	1996 + 1999-2015	1987-2015
Parliament ary careers in compariso n	1949 - now	Yes	No	1949- 2019
Parline	1967-2024	Only aggregate	Only aggregate	1971-2021	1967-2023	1968-2024	1968-2022	1969-2023	1968-2022	1970-2023	1968-2022	1970-2019	1974-2023
Pathways	1990 - 2017	Yes	Yes	.	1993-2016	1993-2012	1992-2013	.	1990-2017	.	.	.	2000-2015
Politicians on Wikipedia and DBpedia	2001 - 2016	Yes	Yes	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016	2001-2016
Unequal Democraci es	2020-2021	Yes	Unknown	.	2020-2021	2020-2021	.	2020-2021
Who Governs	1966 - 2021	Yes	Yes	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021	1966-2021
<u>Regional</u>													
Pathways	1990 - 2017	Yes	Yes	.	1993-2016	1992-2012	1992-2013	.	1990-2017	.	.	.	2000-2015
Vanlangen akker et al. (2013)	1999-2009	Yes	Unknown	.	1999-2006 (Catalonia)	.	.	.	1999-2009 (Saxony)



Table 2 Summary of the variables included in the main comparative datasets on descriptive representation useful for ActEU found through the systematic search. (Carles Pamies & Laura Morales, CSIC)

Database	Variables covered	Citation / URL
<u>European</u>		
Integrity Watch EU	Country, political group, gender, roles in committees, meetings with lobbyists, income, side activities.	Transparency International EU (2023). <i>EU Integrity Watch</i> . https://data.integritywatch.eu/
Lupu and Warner (2002)	Country, year, gender, occupation, income, ideology, party affiliation, social class, and attitudinal variables (for citizens and MPs). Codebook not available.	Lupu, Noam, & Warner, Zach (2022). Affluence and congruence: unequal representation around the world. <i>The Journal of Politics</i> , 84(1), 276-290. https://doi.org/10.1086/714930
Members of the European Parliament	Country, political group, national party, newcomer, re-elected, number of past legislatures in the EP, previous political experience, previous national / regional / local positions, gender, roles in the EP and in committee.	Salvati, Eugenio (2022). <i>Dataset on Members of the European Parliament (1979-2019)</i> [Dataset, version 5.1]. Harvard Dataverse. https://doi.org/10.7910/DVN/V2FJEF
<u>National</u>		
Comparative legislators database	Country, parliament, name, sex, ethnicity, religion, birth, birthplace, party, constituency, offices, occupations, and several others.	Göbel, Sascha & Munzert, Simon (2022). The comparative legislators database. <i>British Journal of Political Science</i> , 52(3), 1398-1408. https://doi.org/10.1017/S0007123420000897 . Also available here: https://complegdatabase.com/
EurElite	Country, year, party, education, age, local political experience, position in government, party employee, gender, occupation, religious denomination, turnover	Best, Heinrich & Edinger, Michael (2005). Converging representative elites in Europe? An introduction to the EurElite project. <i>Czech Sociological Review</i> , 41(3), 499-510. https://www.jstor.org/stable/41132162
Every Politician	Country, name, legislature, social media accounts, Wikipedia/Wikidata identifiers, political group, constituency, gender	MySociety (2018). <i>EveryPolitician</i> . Available from https://everypolitician.org
Global Leadership Project	For multiple leadership positions, including legislators (many variables with missing values for entire countries): country, year, birthday, birthplace, birth country, marital status, occupation, native languages, education, party, party positions, political experience, office types, office names, constituency, previous offices,	Gerring, John, Oncel, Erzen, Morrison, Kevin, and Pemstein, Daniel (2019). Who Rules the World? A Portrait of the Global Leadership Class. <i>Perspectives on Politics</i> , 17(4), 1079-1097. https://doi.org/10.1017/S1537592719000744 . Dataset available here: http://glp.la.utexas.edu/
Gouglas et al. (2018)	Country, year, MP remuneration, and potentially other variables, but technical information not detailed enough. Also contextual level variables such as term duration, party ideology, gender quotas at the party level, legislative turnover...	Gouglas, Athanassios, Bart Maddens and Marleen Brans. 2018. Determinants of legislative turnover in Western Europe, 1945-2015. <i>European Journal of Political Research</i> , 57(3):637-661. https://doi.org/10.1111/1475-6765.12246
InclusiveParl / Repchance	The same as Pathways and: LGBTI+, disability	Not public yet.
Lupu and Warner (2022)	Country, year, gender, occupation, income, ideology, party affiliation, social class, and attitudinal variables (for citizens and MPs). Codebook not available.	Lupu, Noam, & Warner, Zach (2022). Affluence and congruence: unequal representation around the world. <i>The Journal of Politics</i> , 84(1), 276-290. https://doi.org/10.1086/714930



Parliamentary careers in comparison	Multiple variables at the politician career level: country, year, name, gender, birth date, birthplace, citizenship, education, marital status, number of children, occupation, party, electoral and district variables, tenure and office variables.	Bailer, Stefanie, Manow, Philip, & Hug, Simon (2019). <i>Parliamentary Careers in Comparison: Research Project</i> . Basel, Bremen, Geneva: Retrieved from http://parliamentarycareersincomparison.org/
Parline	Country, year, age, contextual factors such as quotas and reserved seats	Inter-Parliamentary Union (2024). PARLINE Database on National Parliaments https://data.ipu.org/
Pathways	Country, year, legislature, name, sex, education, occupation, birth year, family origins, ethnic identity, religious affiliation, identifiable minority, native speaker, party affiliations, constituency, multiple variables on parliamentary tenure, political experience at national, regional, local and EP level, parliamentary roles, committee memberships	Morales, Laura; Vintila, Daniela; Geese, Lucas; Mügge, Liza; van der Pas, Daphne; van de Wardt, Marc (2017). Codebook and data collection guidelines of work package 1 on descriptive political representation in national parliaments of the project Pathways to Power. https://doi.org/10.7910/DVN/MRKJF Harvard Dataverse, V1
Politicians on Wikipedia and DBpedia	Country, name, gender, nationality, occupation, birth date, death date, profession and for many politicians also the political party	Wagner, Claudia (2017). <i>Politicians on Wikipedia and DBpedia</i> . GESIS, Cologne. Data File Version 1.0.0, https://doi.org/10.7802/1515 .
Unequal Democracies	Country, age, gender, occupation, social class, education, party	Elsässer, Lea & Schäffer, Armin (2022). <i>Unequal Democracies: Who does (not) have a seat in parliament?</i> Friedrich Ebert Stiftung. https://library.fes.de/pdf-files/bueros/wien/19349-20220926.pdf
Who Governs	Only for cabinet members: country, year, name, position, gender, birth year, party, cabinet position type, portfolios, prestige.	Nyrup, Jacob, & Bramwell, Stuart (2020). Who governs? A new global dataset on members of cabinets. <i>American Political Science Review</i> , 114(4), 1366-1374. https://doi.org/10.1017/S0003055420000490 . Dataset and documentation available here: https://politicscentre.nuffield.ox.ac.uk/whogov-dataset/download-dataset/
<u>Regional</u>		
Pathways	Country, year, legislature, name, sex, education, occupation, birth year, family origins, ethnic identity, religious affiliation, identifiable minority, native speaker, party affiliations, constituency, multiple variables on parliamentary tenure, political experience at national, regional, local and EP level, parliamentary roles, committee memberships	Morales, Laura; Vintila, Daniela; Geese, Lucas; Mügge, Liza; van der Pas, Daphne; van de Wardt, Marc (2017). <i>Codebook and data collection guidelines of work package 1 on descriptive political representation in regional parliaments of the project Pathways to Power</i> , https://doi.org/10.7910/DVN/KZMOAM . Harvard Dataverse, V1
Vanlangenakker et al. (2013)	Country, age, sex, length of service as MP, reasons for leaving the parliament	Vanlangenakker, Ine, Wauters, Bram, & Maddens, Bart. (2013). Pushed toward the exit? How female MPs leave parliament. <i>Politics & Gender</i> , 9(1), 61-75. https://doi.org/10.1017/S1743923X12000694



2.2 Existing research and datasets on substantive representation in Europe

A systematic search of studies and databases on substantive representation in Europe also returned a considerable number of datasets of uneven content in terms of the nature of the substantive parliamentary representation activities as well as the metadata (variables) on the individual legislators undertaking such parliamentary activities.

We first provide a comprehensive list of the main publications and datasets we considered and then present the key summarising tables of their main attributes.

2.2.1 List of publications and datasets on substantive representation in Europe mapped for the ActEU project

- Baumgartner, Frank R., Christian Breunig & Emiliano Grossman (2019). *Comparative policy agendas: Theory, tools, data*. Oxford University Press. [Note: Comparative Agendas Project].
- Béchara, Hannah, Herzog, Alexander, Jankin, Slava, & John, Peter (2021). Transfer learning for topic labeling: Analysis of the UK House of Commons speeches 1935–2014. *Research & Politics*, 8(2), 20531680211022206. <https://journals.sagepub.com/doi/full/10.1177/20531680211022206>
- Blätte, Andreas (2020). *GermaParl. Download and Augment the Corpus of Plenary Protocols of the German Bundestag*. R package version 1.4.1. <https://polmine.github.io/GermaParl/>
- Blätte, Andreas (2020). *GermaParl. Linguistically Annotated and Indexed Corpus of Plenary Protocols of the German Bundestag*. CWB corpus version 1.0.6. <https://doi.org/10.5281/zenodo.3735141>
- Blätte, Andreas (2018). Zum Verwecheln ähnlich? Eine Klassifikationsanalyse parlamentarischen Diskursverhaltens auf Basis des PolMine-Plenarprotokollkorpus", in: *Computational Social Science*. Die Analyse von Big Data, Nomos 139-162.
- Blätte, Andreas & Leonhardt, Christoph (2020). *MigParl. A Corpus of Speeches on Migration and Integration in Germany's Regional Parliaments*. <https://polmine.github.io/MigParl/>
- Braby, Daniel; Stewart, Fraser (2021). *parlScot: a dataset of 1.8 million spoken contributions from the Scottish Parliament 1999-2021*. <https://doi.org/10.7910/DVN/EQ9WBE>, Harvard Dataverse, V1



- Celis, Karen, Childs, Sarah, & Curtin, Jennifer (2016). Specialised parliamentary bodies and the quality of women's substantive representation: a comparative analysis of Belgium, United Kingdom and New Zealand. *Parliamentary Affairs*, 69(4), 812-829. <https://doi.org/10.1093/pa/gsw007>
- Curini, Luigi, Decadri, Silvia, Ferrara, Alfio, Montanelli, Stefano, Negri, Fefra, & Periti, Francesco (2024). The Gender Gap in Issue Attention and Language Use within a Legislative Setting: An Application to the Italian Parliament (1948–2020). *Politics & Gender*, 20(1), 182-211. <https://doi.org/10.1017/S1743923X2300051X>
- Eggers, Andrew C. & Spirling, Arthur (2014). Electoral security as a determinant of legislator activity, 1832–1918: New data and methods for analyzing British political development. *Legislative Studies Quarterly* 39(4): 593-620. <https://doi.org/10.1111/lsq.12059>
- Erjavec, Tomaž., Kopp, Matyáš, Ljubešić, Nikola, Kuzman, Taja, Rayson, Paul, Osenova, Petya, ... & Fišer, Darja (2024). *ParlaMint II: Advancing Comparable Parliamentary Corpora Across Europe*. <https://doi.org/10.21203/rs.3.rs-4176128/v1>
- Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya, Ljubešić, Nikola, Simov, Kiril, Pančur, Andrej, ... & Fišer, Darja (2023). The ParlaMint corpora of parliamentary proceedings. *Language resources and evaluation*, 57(1), 415-448. <https://doi.org/10.1007/s10579-021-09574-0> [Note: original ParlaMint]
- Grace, Joan (2016). Presence and Purpose in the Canadian House of Commons: The Standing Committee on the Status of Women. *Parliamentary Affairs*, 69(4), 830-844. <https://doi.org/10.1093/pa/gsw008>
- Grace, Joan, & Sawyer, Marian (2016). Representing gender equality: specialised parliamentary bodies. *Parliamentary Affairs*, 69(4), 745-747. <https://doi.org/10.1093/pa/gsw004>
- Haman, Michael & Školník, Milan (2021). Politicians on Social Media. The online database of members of national parliaments on Twitter. *Profesional de la información*, 30(2), e300217. <https://doi.org/10.3145/epi.2021.mar.17>
- Harder, Mette Marie Stæhr (2017). Assessing the impact of parliamentary design: The case of the Danish Committee on Gender Equality. *Scandinavian Political Studies*, 40(4), 434-456. <https://doi.org/10.1111/1467-9477.12102>
- Harder, Mette Marie Stæhr (2023). Parting with 'interests of women': how feminist scholarship on substantive representation could replace 'women's interests' with 'gender equality interests'. *European Journal of Politics and Gender*, 6(3), 377-394. <https://doi.org/10.1332/251510821X16742321076420>



- Herzog, Alexander & Mikhaylov, Slava (2017). *Database of Parliamentary Speeches in Ireland, 1919-2013*. <https://doi.org/10.7910/DVN/6MZN76>, Harvard Dataverse, V2
- Holli, Anne Maria, & Harder, Mette Marie Stæhr (2016). Towards a dual approach: comparing the effects of parliamentary committees on gender equality in Denmark and Finland. *Parliamentary Affairs*, 69(4), 794-811. <https://doi.org/10.1093/pa/gsw006>
- Johnson, N., & Josefsson, C. (2016). A new way of doing politics? Cross-party women's caucuses as critical actors in Uganda and Uruguay. *Parliamentary Affairs*, 69(4), 845-859. <https://doi.org/10.1093/pa/gsw011>
- Kiss, Rebeka & Sebők, Miklós (2022). Creating an enhanced infrastructure of parliamentary archives for better democratic transparency and legislative research: Report on the OPTED forum in the European Parliament (Brussels, Belgium, 15 June 2022). *International Journal of Parliamentary Studies*, 2(2), 278-284. https://brill.com/view/journals/parl/2/2/article-p278_010.xml?language=en
- Koehn, Philipp (2005). EuroParl: A Parallel Corpus for Statistical Machine Translation. In J. Hutchins (Ed.), *The Tenth Machine Translation Summit Proceedings of Conference* (pp. 79-86). International Association for Machine Translation. <https://aclanthology.org/2005.mtsummit-papers.11>
- Leonhardt, Christoph, & Blätte, Andreas (2020). *AustroParl Corpus of Parliamentary Debates* (v.0.1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3819505>
- Leonhardt, Christoph & Blätte, Andreas (2020). *ParisParl Corpus of Parliamentary Debates* (v0.1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3819374>
- Licht, Hauke & Lind, Fabienne (2023). Going cross-lingual: A guide to multilingual text analysis. *Computational Communication Research*, 5(2), 1. <https://doi.org/10.5117/CCR2023.2.2.LICH>
- Mochtak, Michal, Peter Rupnik, and Nikola Ljubešić (2023). The ParlaSent multilingual training dataset for sentiment identification in parliamentary proceedings. *arXiv preprint arXiv:2309.09783*. <https://doi.org/10.48550/arXiv.2309.09783>
- PolMine (2022). *Registry of Research Data Repositories*. <http://doi.org/10.17616/R3Z36S>
- Raiber, Klara & Spierings, Niels (2022). An agnostic approach to gender patterns in parliamentary speech: a question of representation by topic and style. *European Journal of Politics and Gender*, 5(3), 361-381. <https://t.co/UeHxxnOGtd>



- Rangoni, Sacha, Bedock, Camille, & Talukder, David (2023). More competent thus more legitimate? MPs' discourses on deliberative mini-publics. *Acta Politica*, 58(3), 531-551. <https://link.springer.com/article/10.1057/s41269-021-00209-4>
- Rauh, Christian & Schwalbach, Jan (2020). *The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies*. <https://doi.org/10.7910/DVN/L4OAKN>, Harvard Dataverse, V1
- Sawer, Marian (2020). Gender mainstreaming and the substantive representation of women: Where do parliamentary bodies fit? *Politics, Groups, and Identities*, 8(3), 648-660. <https://doi.org/10.1080/21565503.2020.1752444>
- Sawer, Marian & Turner, Alicia (2016). Specialised parliamentary bodies: their role and relevance to women's movement repertoire. *Parliamentary Affairs*, 69(4), 763-777. <https://doi.org/10.1093/pa/gsw003>
- Schwemmer, Carsten (2021). *The Pathways Corpus - data of work package 2 on substantive political representation*. <https://doi.org/10.7910/DVN/JMD0KM>, Harvard Dataverse, V1.
- Sebők, Miklós, Proksch, Sven-Oliver, Rauh, Christian, Visnovitz, Péter, Balázs, Gergő & Schwalbach, Jan (2023). Comparative European legislative research in the age of large-scale computational text analysis: A review article. *International Political Science Review*. <https://doi.org/10.1177/01925121231199904> [Note: output linked to the OPTED project]
- Sebők, Miklós, Proksch, Sven-Oliver & Rauh, Christian (2021). *OPTED. Review of Available Parliamentary Corpora*. https://opted.eu/fileadmin/user_upload/k_opted/OPTED_Deliverable_D5.1.pdf
- Talukder, David (2023). Challenging assumptions: investigating measurement sensitivity in substantive representation and its effects on policy satisfaction. *European Journal of Politics and Gender*, 6(3), 414-432. <https://doi.org/10.1332/251510821X16889871242153>
- van Aggelen, Astrid, Laura Hollink, Max Kemman, Martijn Kleppe, and Henri Beunders (2017). The debates of the European Parliament as linked open data. *Semantic Web*, 8, no. 2: 271-281. [LinkedEP output regarding the project, data models and data creation methods] <https://doi.org/10.3233/SW-160227>
- Wäckerle, Jens (2023). A Bill of their Own: Collaboration of Women in European Parliaments. *European Journal of Politics and Gender*, 6(3), <https://doi.org/10.1332/251510821X16782058189028>



- Wäckerle, Jens & Castanho Silva, Bruno. 2023. Distinctive Voices: Speech and Women's Representation in Five European Parliaments. *Legislative Studies Quarterly, Early View*. Available here (chapter 4): https://kups.ub.uni-koeln.de/61756/1/PhD_Dissertation_Waeckerle.pdf
- Wängnerud, Lena (2009). Women in parliaments: Descriptive and substantive representation. *Annual Review of Political Science*, 12(1), 51-69. <https://doi.org/10.1146/annurev.polisci.11.053106.123839>

2.2.2 Summary of key attributes of the main comparative datasets on substantive representation in Europe mapped for the ActEU project

After this mapping exercise, we have selected the main datasets that provided comparative corpora on parliamentary activities and texts across several countries in Europe and which could be of use for the ActEU project as sources for substantive representation by legislators, political parties and legislative assemblies as a whole. Unlike the [OPTED project](#) (funded by Horizon 2020), our aim is not to produce a comprehensive inventory, but to focus on the availability of processed datasets that could be useful for the ActEU project.

Hence, we have focused on datasets providing access to parliamentary corpora that are already processed and contain at least some minimal metadata / variables about the individual legislators who were generating the text / speech (see Table 3). At this initial stage, we have been inclusive and recorded the main databases regardless of the type of parliamentary initiative or type of intervention it focuses on. Nevertheless, as Table 4 shows, the majority of the selected datasets present corpora that are generated from plenary debates or speeches, or written questions submitted to the executive. The Comparative Agendas Project is the only one providing a wider range of corpora.

When assessing the metadata / variables that these corpora provide at the legislator level, the main conclusion is that the majority of the available corpora provide very limited leverage to analyse the connection between descriptive and substantive representation, as only very few attributes of the individual legislators generating the speech / text are typically provided: gender, age and political party for the most part.

Given these results of our stock-taking and mapping exercise, the teams working on Tasks 4.1 and 4.2 of WP4 and involved in this deliverable 4.1 (CSIC, PLUS, AUTH), decided to take as the starting point for the datasets that would be shared within the ActEU team the structure and methodological decisions taken in the datasets produced by the [InclusiveParl project](#) led by Laura Morales (CSIC, previously Sciences Po). This provided the baseline for robust indicators at the legislator individual-level that would already cover four countries considered in ActEU: France, Italy, Spain and the United Kingdom. This choice facilitated the extension of data for a reasonable amount of legislator-level variables to two cases more: Austria and Greece. The resulting datasets are described in sections three and four of this report.



Table 3 Summary of datasets on substantive representation found through the systematic search. (Carles Pamies & Laura Morales, CSIC)

Database	Period covered	CZ	ES	FR	IT	PL	DE	FI	DK	AT	GR
<u>European</u>											
LinkedEP	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017	1999-2017
<u>National</u>											
Comparative Agendas Project	1953-2018	.	1977-2018	(No Qs, but legislation)	1996-2014	.	1976-2005
Curini et al. (2024)	1948-2020	.	.	.	1948-2020
OPTED inventory	1267-2021	1920-2021	1977-2021	1979-2021	1983-2021	1919-2021	1949-2021	1734-2019	1665-2021	1753-2021	2010-2019
ParlaMint I/II (CLARIN)	1977-2023	1993-2023	1996-2023	2017-2022	2013-2022	2015-2022	.	.	1997-2022	1996-2022	.
ParlSpeech v2	1977 - 2023	1993-2023	1996-2023	2017-2022	2013-2022	2015-2022	1991-2018	.	1997-2018	1996-2018	.
Pathways	1 legislative term in the mid -2010s		2011-2015	2007-2012	2008-2013		2009-2013				2015
PolMine	1996-2019	.	.	1996-2019	.	.	1996-2016	.	.	1996-2019	.
<u>Regional</u>											
ParlaMint II	2015-2022	2015-2022	2015-2022	2015-2022	2015-2022	2015-2022	.	2015-2022	2015-2022	2015-2022	2015-2022
MigParl (part of PolMine)	2000-2018	2000-2018



Table 4 Summary of the variables included in the main comparative datasets on substantive representation useful for ActEU found through the systematic search. (Carles Pamies & Laura Morales, CSIC)

Database	Variables of interest for the ActEU project	Description	Citation / URL
European			
LinkedEP		Type of data: Plenary debates of the European Parliament	van Aggelen, Astrid, Laura Hollink, Max Kemman, Martijn Kleppe, and Henri Beunders (2017). The debates of the European Parliament as linked open data. <i>Semantic Web</i> , 8, no. 2: 271-281. [LinkedEP output regarding the project, data models and data creation methods] https://doi.org/10.3233/SW-160227
National			
Comparative Agendas Project	Many, depending on the specific study	Type of data: bills, oral and written questions, legislation, debates, manifestos...	Baumgartner, Frank R., Christian Breunig & Emiliano Grossman (2019). <i>Comparative policy agendas: Theory, tools, data</i> . Oxford University Press.
Curini et al.(2024)	Gender + committee, seniority, party ideology, topics	Type of data: parliamentary speeches	Curini, Luigi, Decadri, Silvia, Ferrara, Alfio, Montanelli, Stefano, Negri, Fefra, & Periti, Francesco (2024). The Gender Gap in Issue Attention and Language Use within a Legislative Setting: An Application to the Italian Parliament (1948–2020). <i>Politics & Gender</i> , 20(1), 182-211. https://doi.org/10.1017/S1743923X2300051X
OPTED	Gender.av (gender) + party.av (party)	OPTED provides an inventory and access to political texts that have been produced by other projects/researchers/sources. Type of data: Legislative speeches, questions and legislative documents	Kiss, Rebeka, & Sebők, Miklós (2022). Creating an enhanced infrastructure of parliamentary archives for better democratic transparency and legislative research: Report on the OPTED forum in the European Parliament (Brussels, Belgium, 15 June 2022). <i>International Journal of Parliamentary Studies</i> , 2(2), 278-284. https://brill.com/view/journals/parl/2/2/article-p278_010.xml?language=en
ParlaMint I/II (CLARIN)	Gender (gender) + Birth (year of birth)	Type of data: Plenary speeches in national parliaments	Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya, Ljubešić, Nikola, Simov, Kiril, Pančur, Andrej, ... & Fišer, Darja (2023). The ParlaMint corpora of parliamentary proceedings. <i>Language resources and evaluation</i> , 57(1), 415-448. https://doi.org/10.1007/s10579-021-09574-0 Erjavec, Tomaž., Kopp, Matyáš, Ljubešić, Nikola, Kuzman, Taja, Rayson, Paul, Osenova, Petya, ... & Fišer, Darja (2024). <i>ParlaMint II: Advancing Comparable Parliamentary Corpora Across Europe</i> . https://doi.org/10.21203/rs.3.rs-4176128/v1
ParlSpeech v2		Type of data: Parliamentary speeches	Rauh, Christian & Schwalbach, Jan (2020). <i>The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key</i>



			<i>legislative chambers of nine representative democracies.</i> https://doi.org/10.7910/DVN/L4OAKN , Harvard Dataverse, V1
Pathways	All in descriptive representation dataset can be merged	Type of data: written questions to the Executive	Schwemmer, Carsten (2021). <i>The Pathways Corpus - data of work package 2 on substantive political representation.</i> https://doi.org/10.7910/DVN/JMDOK , Harvard Dataverse, V1.
PolMine	Speaker, party affiliation, parliamentary group affiliation, role, legislative period, session, date, interjections, year and agenda item	Type of data: plenary debates, speeches in parliament	PolMine (2022). <i>Registry of Research Data Repositories.</i> http://doi.org/10.17616/R3Z36S . [Note: The PolMine project includes corpora from the AustroParl, GermaParl, ParisParl, MigParl and United Nations General Debate Corpus projects]
Regional			
Parlamint II	Gender (gender) + Birth (year of birth)	Type of data: Plenary speeches in regional parliaments	Erjavec, Tomaž., Kopp, Matyáš, Ljubešić, Nikola, Kuzman, Taja, Rayson, Paul, Osenova, Petya, ... & Fišer, Darja (2024). <i>ParlaMint II: Advancing Comparable Parliamentary Corpora Across Europe.</i> https://doi.org/10.21203/rs.3.rs-4176128/v1
MigParl	Session, agenda, date, speaker, role of speaker, party, regional state where debate held, sum probability for migration and integration relevance	Type of data: speeches on migration and integration affairs in Germany's regional parliaments	Blätte, Andreas and Leonhardt, Christoph (2020). <i>MigParl. A Corpus of Speeches on Migration and Integration in Germany's Regional Parliaments.</i> https://polmine.github.io/MigParl/

3 The ActEU datasets on descriptive political representation

The ActEU project members and tasks require information about descriptive political representation (presence, who 'stands for' citizens) that can be used as external sources of indicators on the degree to which the parliaments (and the parties with seats in these parliamentary chambers) of the ten countries studied resemble or mirror the diversity of the populations of their respective countries. To this end, we have generated databases that aggregate at the party and country-level information on the number of legislators with a range of socio-demographic characteristics that allow gauging the diversity of descriptive representation.

Providing the datasets for the 10 countries studied in ActEU was not feasible, given the person-month resources and timing of this deliverable when considered jointly with the timing of other deliverables where the partner teams are involved, but we provide at this stage datasets for 6 European countries (Austria, France, Greece, Italy, Spain and the United Kingdom) and new datasets will be provided for the remaining countries (Czechia, Denmark, Finland, Germany and



Poland) at a later stage. The key difference between the datasets from the 6 countries that constitute the current delivery and the other 5 will be the data sources and mode of data collection. Whereas the data collected for the former are all based on the InclusiveParl methodology and codebooks, the data that will be shared for the latter will be compiled from a range of public data sources presented in Tables 1 and 2 in this report. The reason for this is that the teams involved in these tasks of WP4 lack the linguistic skills and country expertise to collect primary data about legislators for Czechia, Denmark, Finland, Germany and Poland.

3.1 Data collected, methodology and variables

The datasets generated by the three partner teams involved in Tasks 4.1 and 4.2 and contributing to Deliverable 4.1 include information for 12 legislative terms across 6 European countries for nearly 5,900 legislators (see Table 5). The datasets that are shared in the ActEU Sciebo internal cloud and the [ActEU community Zenodo](#) repository include 26 variables aggregated at the country/legislative term/party level (see Table 6).

The main aim was to cover, for all the countries for which this was feasible, the information on sex and gender identity, age, education level, country of birth, nationality at birth, family ancestry, identifiable minority status, known LGBTI+ identity and whether they were newly elected for all legislators who took up a seat in the lower chambers or unicameral assemblies for the legislative terms identified in Table 5.

Data collection was undertaken through manual coding, triangulating multiple publicly available biographical sources about the respective legislators by each of the respective teams who took responsibility for the datasets they would generate (see the full description of the methodological approach in Morales et al. 2017 and Vincent-Mory and Morales 2024).

Table 5 Description of the data collected (Laura Morales and Carles Pamies, CSIC)

Country	Legislative term	Number of legislators	Team that collected the data
Austria	27 (2019-2024) &	213	PLUS
	26 (2017-2019)	210	
France	15 (2017-2022)	656	Sciences Po, CSIC
Greece	18 (2019-2023)	309	AUTH
Italy	18 (2018-2022) &	660	Sciences Po, CSIC
	17 (2013-2018)	672	
Spain	14 (2019-2023), 13	407	Sciences Po, CSIC
	(2019-2019) & 12	355	
	(2016-2019)	393	
United Kingdom	58 (2019-2024), 57	673	Sciences Po, CSIC
	(2017-2019) &	656	
	56 (2015-2017)	660	



Table 6 Description of the variables included in the descriptive representation dataset (Laura Morales and Carles Pamies, CSIC)

VARIABLE	CONTENT	VALUES
country_id_NUTS	NUTS code for the country	Only standard NUTS codes for the country
LegislatNum	Number of the legislature (as given by Parliament) in Arabic numbers (do not use Roman numbers)	E.g. 12, 13, 14, etc
LegislatBegin	Date in which the legislature begins: YYYY-MM-DD	E.g., 2019-11-03 (for 3rd Nov 2019)
LegislatEnd	Date in which the legislature ends: YYYY-MM-DD	E.g., 2023-06-20 (for 20th June 2023)
PartyName	Full name of party or list for the aggregate values in each row	In original language but Western European alphabets (no Greek characters, please)
PartyAcronym	Acronym of party or list for the aggregate values in each row	The acronym used by the party or assigned in Appendix 2
PartyCodeCMP	Party code according to the CMP dataset	Use exactly the CMP code
party_id	Pathways/ActEU identifier for party	Pathways identifier for party in Appendix 2 (for countries covered by Pathways) or new unique code by party for new ActEU countries
N_total_MPs	Number of MPs that took up a seat with this party during this term	A numerical value (if not known, assign -99)
N_Women	Number of women MPs that took up a seat with this party during this term	A numerical value (if not known, assign -99)
N_Men	Number of men MPs that took up a seat with this party during this term	A numerical value (if not known, assign -99)
N_NonBinary	Number of 'out' non-binary MPs that took up a seat with this party during this term	A numerical value (if not known, assign -99)
N_Trans	Number of 'out' trans MPs that took up a seat with this party during this term	A numerical value (if not known, assign -99)
N_Youth_up29	Number of MPs who were 29 yrs-old or younger at the time of taking up a seat with this party during this term	A numerical value (if not known, assign -99)
N_Edu=UnivDegree	Number of MPs who had a university degree at the time of taking up a seat with this party during this term	A numerical value (if not known, assign -99)



N_Rookie=1	Number of MPs who took up a seat with this party during this term that were 'rookies' (i.e., took up for the first time a seat in the lower chamber)	A numerical value (if not known, assign -99)
N_BornC=0	Number of MPs who took up a seat with this party during this term that were born in a country other than the country coded	A numerical value (if not known, assign -99)
N_BornC=1	Number of MPs who took up a seat with this party during this term that were born in the country coded	A numerical value (if not known, assign -99)
N_BornC=-99	Number of MPs who took up a seat with this party during this term for which the country where they were born is not known	A numerical value (if not known, assign -99)
N_NationalC=0	Number of MPs who took up a seat with this party during this term that were NOT born with the nationality of the country coded	A numerical value (if not known, assign -99)
N_NationalC=1	Number of MPs who took up a seat with this party during this term that were born with the nationality of the country coded	A numerical value (if not known, assign -99)
N_NationalC=-99	Number of MPs who took up a seat with this party during this term for which the country of nationality with which they were born is not known	A numerical value (if not known, assign -99)
N_CIO2bis=1	Number of MPs who took up a seat with this party during this term for which the Stata recoded variable CIO2bis is 1 (value 'yes' for "MP of immigrant origin")	A numerical value (if not known, assign -99). NOTE: This variable can only be created from the individual-level data following the Pathways/InclusiveParl Stata do file. If you have not collected the full individual-level data, you will need to code -99 for all parties.
N_IdentifiableMinority=1	Number of MPs who took up a seat with this party during this term that are members of an identifiable minority	A numerical value (if not known, assign -99)
N_LGBTI=1	Number of MPs who took up a seat with this party during this term that openly identify as LGBTI+	A numerical value (if not known, assign -99)
DataSources	The data sources used to compile the data for this party/term row	A string variable where you describe the sources

Note: The extended definition and operationalization of all these variables can be found in Morales et al. (2017) and Vincent-Mory and Morales (2024), see section A1 in the Appendix for the full citation and DOI.



3.2 Descriptive statistics of the datasets collected

Table 7 provides the descriptive statistics for some of the key variables included in the datasets on biographical information of legislators in the six countries studies. These statistics are calculated at the country/legislature/party level aggregating for each country all the legislative terms covered. In other words, if for a given country (e.g., Spain and the UK) the period considered covers three legislative terms, the statistics provide the minimum, maximum and average values at the party level for the pooled dataset of the three legislative terms.

Table 7 Descriptive statistics of the key variables included in the descriptive representation datasets (Laura Morales and Carles Pamies, CSIC)

Variable	Austria	France	Greece	Italy	Spain	United Kingdom
N_Women	[1 - 32]; 15	[0 - 163]; 10	[2 - 26]; 11	[0 - 120]; 15	[0 - 64]; 7	[0 - 124]; 19
N_Men	[1 - 50]; 24	[0 - 178]; 15	[4 - 134]; 40	[0 - 200]; 33	[0 - 93]; 9	[0 - 279]; 38
N_NonBinary	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0
N_Trans	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 0]; 0	[0 - 1]; 0
N_Youth_up29	[0 - 7]; 3	[0 - 16]; 1	[0 - 1]; 0	[0 - 28]; 3	[0 - 5]; 0	[0 - 8]; 1
N_BornC=0	[0 - 5]; 2	[0 - 22]; 1	[1 - 2]; 2	[0 - 4]; 1	[0 - 5]; 0	[0 - 18]; 3
N_NationalC=0	Not provided	[0 - 20]; 1	[0 - 0]; 0	[0 - 3]; 0	[0 - 2]; 0	[0 - 10]; 2
N_IdentifiableMinority=1	[0 - 10]; 6	[0 - 42]; 2	[0 - 1]; 0	[0 - 4]; 1	[0 - 3]; 0	[0 - 44]; 5
N_LGBTI=1	[0 - 4]; 1	[0 - 7]; 1	[0 - 0]; 0	[0 - 2]; 0	[0 - 5]; 1	[0 - 24]; 4
Total N MPs	[2 - 82]; 39	[1 - 341]; 25	[6 - 160]; 51	[1 - 320]; 48	[1 - 157]; 15	[1 - 369]; 57

Note: For each country we show the descriptive statistics as [minimum value - maximum value]; mean value. These statistics are calculated at the party level for each country for the given legislative term and have been rounded down or up to the closest integer value.

Sources: Gianna, Ceron and Lefkofridi (2024), Gotinakos and Teperoglou (2024), Morales and Pamies (2024a, 2024b, 2024c, 2024d).

As we can see, the datasets will provide us with considerable cross-national variation regarding the descriptive representation of core socio-demographic groups and categories for a range of countries that vary considerably in their electoral systems and the proportionality of the seats allocated to parties compared to the share of votes they received. Such detailed analyses will be included in Deliverable 4.2.



4 The ActEU dataset on substantive political representation

After consideration of the existing datasets and their suitability to match as much as possible parliamentary activities with the individual legislators' attributes, the teams collaborating in Tasks 4.1 and 4.2 decided to focus, for these tasks, on written questions submitted by legislators to the executive branch of the government of their respective countries. Written questions have the advantage that they can more easily be initiated by individual legislators out of their own interest in representing specific topics, territories, issues or groups, as they are less subject to strict parliamentary group leadership control and assignment (see, e.g., Martin and Rozenberg 2014, Höhmann and Sieberer 2020, Rozenberg and Martin 2020, Randour et al. 2022, De Vet and Devroue 2023). Thus, while still reflecting broadly ideological party lines, written questions allow us to approximate the underlying issue and demand representation preferences of individual legislators.

4.1 Data collected, methodology and variables

We have aimed to collect the written questions of the most recent legislative term for each country, where this was feasible (see Table 8). Written questions were collected through Open Data APIs for Austria, France, Italy and the United Kingdom; whereas they required data requests for Spain. While the text questions for the Austrian parliament were available in open data files (in JSON format), they required a two-stage retrieval process that was time consuming.

In the case of Greece, full access to the written questions has not been obtained as, on the one hand, the Hellenic Parliament does not have these types of parliamentary initiatives available for direct download through the Open Data API and, on the other hand, web-scraping access was initially blocked. We contacted directly the Hellenic Parliament IT and research staff and we were, furthermore, informed that the available information on the normal search motor on their website was not fully complete. Indeed, for the 2019-2023 parliamentary term, the public search motor provided by the Hellenic Parliament yields a combined result of 29,483 written questions submitted to the executive.⁴ Instead, the parliamentary services have communicated that their internal databases suggest that a total of 47,195 written questions were submitted by legislators to the executive. Hence, the discrepancy being too large (approximately 40% of the written questions were missing from the freely accessible sources), we decided not to include the Greek corpus for the time being. A collaboration with the Hellenic Parliament is possible and is being studied but cannot be set up within the deadline for the current report.

A second challenge we faced concerned the Spanish written questions. In this case, the Spanish parliamentary website of Congreso de los Diputados provided only in ready-to-use machine readable format the key metadata: a title/heading for the question summarising the main content of the question, the date when it was submitted, the name and surnames of the legislator(s) who

⁴ The search motor is retrievable here: <https://www.hellenicparliament.gr/Koinovouleftikos-Elenchos/Mesa-Koinovouleutikou-Elegxou?subject=&protocol=&type=63c1d403-0d19-409f-bb0d-055e01e1487c&SessionPeriod=d6ba96d9-b6b6-44cd-86e1-af2300d60d6f&partyId=&mpId=&ministry=&datefrom=&dateto=>



authored the question, their parliamentary group of affiliation, the procedural stage at which the question was at the end of the parliamentary term and the links to the PDF files of the full text of the question itself and, when available, of the reply provided by the executive. Besides the fact that the parliamentary website does not provide an API or any other form to download all these metadata and documents in bulk, the main problem faced was that the full texts of the questions were all stored in PDF files that needed to be processed for text extraction. Moreover, for some questions the text was stored in the PDF file with the scanned image of the original letter of the question submitted by the legislator(s) but in other cases the question texts were only included in the official bulletin of parliament (Boletín Oficial de las Cortes Generales, BOCG) embedded in a larger document with agreements passed by the governance body of the chamber (Mesa del Congreso). This latter format of making the questions publicly available requires multiple stages of PDF text extraction, optical character recognition (OCR) and text processing that require more time than was available before this report was due. Although we contacted the parliamentary IT and archive services, it was communicated to us that these services do not have the full text of the questions in any other format than the PDFs. As a consequence, the first version of the data generated with this deliverable only includes data and analyses from the titles/headings for the Spanish written questions. Future version releases will include data and analyses for the full texts of the questions.

All processes of corpus extraction, processing and analysis were undertaken with Python code specifically developed for these tasks and open-source Python libraries. This includes the extraction of metadata and texts from Excel or JSON files, requests to parliament websites when needed, matching written questions and names of the individual legislators with biographical and party affiliation data using Levenshtein distance⁵, tokenization and lemmatization of question texts and keyword dictionaries with spaCy language-specific models⁶, stopword extraction from tf-idf lemmas with scikit-learn software⁷ and keyword counting for dictionary-based analyses using spaCy Matcher⁸.

Table 8 Description of the data collected on substantive representation (Laura Morales and Nayla Escribano, CSIC)

Country	Legislative term	Number of written questions	Team that collected the data
Austria	27 (2019-2024)	19,524	Sciences Po, CSIC
France	15 (2017-2022)	45,665	Sciences Po, CSIC
Greece	18 (2019-2023)*	47,195	Sciences Po, CSIC
Italy	17 (2013-2018)	19,022	Sciences Po, CSIC
Spain	14 (2019-2023) ⁺	108,045	Sciences Po, CSIC
United Kingdom	57 (2017-2019)	107,513	Sciences Po, CSIC

* Not included in version 1 of the datasets. ⁺ Analyses only based on question titles/headings for version 1.

⁵ Software used for calculating Levenshtein distance between legislator names in obtained data and in biographical data: <https://pypi.org/project/Levenshtein/>.

⁶ Language-specific models for tokenization and lemmatization: <https://spacy.io/usage/models>.

⁷ Software used to obtain stopwords from tf-idf lemmas: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.

⁸ Software used to count dictionary keywords: <https://spacy.io/api/matcher/>.



Once the full set of written questions were extracted, we applied topic-based dictionaries that were prepared by the teams contributing to Tasks 4.1 and 4.2. We created nine keyword dictionaries for the core interests of this project: women's issues, men's issues, Trans issues, non-binary persons' issues, LGBTI+ issues, youth issues, climate change issues, immigration issues and minority issues (see Morales et al. 2024). The dictionaries were collectively produced through various stages with the aim of generating keyword dictionaries that would be reasonably comprehensive of the most important terms pointing to issues or topics of specific concern relating to each social category or issue area but without aiming at exhaustivity. The goal was to produce dictionaries that would allow for a reasonable comparison of topic/issue salience across the countries studied, focusing on sub-issues and topics that could reasonably be the subject of attention from legislators in all of the countries without creating country-specific dictionaries. A first version of the lists of keywords produced in English was expanded in two more rounds of review by team members also exclusively for the English version (see Table 11 in section A.3 in the Appendix). Once the lists of keywords were stabilised, they were automatically translated into each of the other five languages with DeepL Translator free service⁹. The automated translations were then reviewed, and where necessary corrected, by the native/proficient speaker members of the team for each language. Whenever disagreements arose, a process of exchange and adjudication was put in place. Following this dictionary preparation stage, keywords for each language were processed as indicated above for tokenization, lemmatization and counting on all country datasets to generate the variables described in Table 9 at the country/legislature/party level.

Table 9 Description of the variables included in the substantive representation datasets (Laura Morales and Nayla Escribano, CSIC)

VARIABLE	CONTENT	VALUES
country_id_NUTS	NUTS code for the country	Only standard NUTS codes for the country
LegislatNum	Number of the legislature (as given by Parliament) in Arabic numbers (do not use Roman numbers)	E.g. 12, 13, 14, etc
LegislatBegin	Date in which the legislature begins: YYYY-MM-DD	E.g., 2019-11-03 (for 3rd Nov 2019)
LegislatEnd	Date in which the legislature ends: YYYY-MM-DD	E.g., 2023-06-20 (for 20th June 2023)
PartyName	Full name of party or list for the aggregate values in each row	In original language but Western European alphabets (no Greek characters, please)
PartyAcronym	Acronym of party or list for the aggregate values in each row	The acronym used by the party or assigned in Appendix 2
PartyCodeCMP	Party code according to the CMP dataset	Use exactly the CMP code
party_id	Pathways/ActEU identifier for party	Pathways identifier for party in Appendix 2 (for countries covered by Pathways) or new unique

⁹ DeepL Translator free service: <https://www.deepl.com/es/translator>.



		code by party for new ActEU countries
N_total_Qs	Total number of questions asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_WomenDict	Total number of questions with terms included in the Women's Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_MenDict	Total number of questions with terms included in the Men's Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_NonBinaryDict	Total number of questions with terms included in the Non-Binary Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_TransDict	Total number of questions with terms included in the Trans Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_YouthDict	Total number of questions with terms included in the Youth Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_ImmigrDict	Total number of questions with terms included in the Immigration Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_MinoritiesDict	Total number of questions with terms included in the Minorities Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_LGBTIDict	Total number of questions with terms included in the LGBTI Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
N_Qs_ClimateDict	Total number of questions with terms included in the Climate Dictionary asked by any MP affiliated to this party at the time of submitting the question in this legislature	A numerical value (if not known, assign -99)
DataSources	The data sources used to compile the data for this party/term row	A string variable where you describe the sources



4.2 Descriptive statistics of the datasets collected

Table 10 shows the descriptive statistics of the resulting variables that capture the salience of each (sub)issue topic based on the dictionaries generated. For the Spanish dataset, they are based on the headings and not on the full texts of the written questions submitted by the legislators. A test undertaken with the Austrian written questions, for which we have both a heading and the full text processed, suggest that using the headings only can considerably underestimate the counts. Although the headings provided in the Austrian open data are considerably shorter than those provided in the Spanish datasets with written questions, these tests suggest that the descriptive statistics in the table below should not be used yet to compare salience across the six countries for this first version of the datasets. The descriptive statistics are provided to illustrate the nature of the data collected.

Table 10 Descriptive statistics of the variables included in the substantive representation datasets (Laura Morales and Nayla Escribano, CSIC)

Variable	Austria	France	Italy	Spain*	United Kingdom
N_total_Qs	[240 - 7507]; 3254	[13 - 17762]; 1903	[2 - 6660]; 1463	[8 - 52281]; 3859	[231 - 71244]; 13439
N_Qs_WomenDict	[6 - 431]; 105	[1 - 1464]; 149	[1 - 437]; 107	[0 - 1777]; 122	[3 - 2988]; 518
N_Qs_MenDict	[3 - 167]; 46	[0 - 903]; 88	[0 - 747]; 158	[0 - 746]; 35	[0 - 351]; 61
N_Qs_NonBinaryDict	[0 - 0]; 0	[0 - 1]; 0	[0 - 70]; 13	[0 - 1]; 0	[0 - 10]; 1
N_Qs_TransDict	[0 - 1]; 0	[0 - 16]; 2	[0 - 13]; 2	[0 - 50]; 2	[0 - 67]; 13
N_Qs_LGBTIDict	[0 - 45]; 8	[0 - 96]; 8	[0 - 58]; 17	[0 - 167]; 7	[1 - 218]; 36
N_Qs_YouthDict	[9 - 438]; 109	[1 - 3232]; 340	[0 - 1021]; 246	[0 - 1882]; 108	[0 - 4417]; 760
N_Qs_ImmigrDict	[0 - 187]; 42	[1 - 1066]; 111	[0 - 487]; 126	[0 - 1241]; 61	[1 - 2425]; 426
N_Qs_MinoritiesDict	[0 - 45]; 16	[0 - 63]; 8	[0 - 60]; 12	[0 - 12]; 2	[0 - 112]; 18
N_Qs_ClimateDict	[10 - 473]; 96	[1 - 2574]; 249	[0 - 2440]; 421	[0 - 1793]; 113	[3 - 4029]; 853

Note: For each country we show the descriptive statistics as [minimum value - maximum value]; mean value. These statistics are calculated at the party level for each country for the given legislative term and have been rounded down or up to the closest integer value.

Source: Escribano and Morales (2024).

* The statistics calculated for the Spanish written questions for this version reflect the counts based exclusively on the heading/title of the question.

Although interpretation of the findings is not the purpose of this report and these will be examined at length in Deliverable 4.2, we can already detect that some issues elicit very little attention despite the very large number of questions submitted by legislators to their respective executives. This is particularly the case of issues relating to LGBTI+ people and concerns. If one focuses on the within-country average salience of each of the (sub)issues and compares them across the country, we can also already detect relevant differences. The findings suggest a greater attention to youth, women and climate change issues in Austria, France, Spain and the United Kingdom, but an apparent greater focus on men's issues in Italy. Nevertheless, the full analysis of the data will be developed in a future report (D4.2).



5 Conclusions and policy-relevant recommendations

This report aimed at presenting and describing in detail the approach, sources, methodology, challenges and contents of the datasets on descriptive and substantive representation produced within WP4 of the ActEU project. We have outlined the scoping and mapping exercises undertaken by the leading team (Sciences Po / CSIC), the data collected and the main attributes of the contents of the datasets. The appendix includes the data citation and references for the datasets generated, as well as for the codebooks and other documentation employed in data production.

This report and deliverable was not designed to produce policy-relevant recommendations but, when assessing the nature and quality of the available datasets, we can share some general recommendations for those in positions to shape policies regarding parliamentary data.

First, it is clear that there is a need for better data from official parliamentary sources that can be augmented and improved by social scientists and computational linguists. Although some parliaments keep good Open Data tools that allow for relatively easy and smooth data access, it is more often the case that national parliaments do not provide Open Data tools that are in line with best practices and standards. There is considerable room for improvement in this area.

Second, the data provided by parliaments on parliamentary activities in Open Data access is often very fragmentary and sometimes incomplete compared to information that can be retrieved through other means (sometimes their ordinary websites). There is a need to set up data maintenance and curation processes that are robust and ensure quality checks. This requires properly funding Open Data teams within the IT services of the national parliaments and integrating the needs of Open Data sharing in parliamentary procedures. For example, the way that legislators submit documentation and initiatives should be fully electronic, such that complete texts are available and processed with limited post-processing by Open Data services.

Third, reliable, complete high-quality data on key socio-demographic characteristics and identities of European Parliament, national and regional legislators is lacking. It would be useful for both transparency and better representation of citizens' expectations to provide more comprehensive biographical data on parliamentary websites. This can be done while respecting the rights to privacy of legislators. For example, more comprehensive biographical forms could be designed where legislators can choose which pieces of information they do not want to share, which ones they wish to share only for research purposes under strict confidentiality protocols, which ones they wish to share publicly only in anonymised form for aggregate analyses and which ones they are happy to share publicly with no restrictions.

In this regard, it would be useful for DG Research, the European Parliament, Eurostat and other relevant bodies at EU level to foster a working group bringing together the main ERICs that could contribute to improve parliamentary data on both legislators' biographical data and their representative activities (CLARIN, CESSDA and DARIAH), a range of academic experts as well as national and (where relevant) regional assemblies. Such a working group could provide good practice standards on the sharing of Open Data about both legislators biographical details as well as parliamentary activities and the many texts and outputs that parliaments generate.



Bibliography/List of References

Bedock, Camille, and Sophie Panel. "Conceptions of democracy, political representation and socio-economic well-being: explaining how French citizens assess the degree of democracy of their regime." *French Politics* 15, no. 4 (2017): 389-417.

Carstens, Jens. "Unpacking political trust: a review of conceptualisations, determinants, and consequences." *French Politics* 21, no. 3 (2023): 295-314.

Christensen, Henrik Serup, Janette Huttunen, Wit Hubert, Mikołaj Cześnik, Alexandros-Christos Gkotlinakos, Eftichia Teperoglou, Jens Carstens, Laura Morales, Carles Pamies, and Chiao Li. 'Report on the General Dynamics of Trust Based on Secondary Data Analysis'. Zenodo, 25 September 2024. <https://doi.org/10.5281/zenodo.13836873>

De Vet, Benjamin, and Robin Devroe. "Party Control, Intraparty Competition, and the Substantive Focus of Women's Parliamentary Questions: Evidence from Belgium." *Politics & Gender* 19, no. 1 (2023): 247-271.

Escribano, Nayla, and Laura Morales. 'Aggregate Dataset on Substantive Representation in the Austrian, British, French, Italian and Spanish Parliaments (2013-2024)'. CSIC, 30 September 2024. <https://doi.org/10.5281/zenodo.13862300>

Gianna, Ermela, Matilde Ceron, and Zoe Lefkofridi. 'Aggregate Dataset on Descriptive Representation in the Austrian Parliament (2017-2024)'. PLUS, 30 September 2024. <https://doi.org/10.5281/zenodo.13861846>

Gotinakos, Alexandros-Christos, and Eftichia Teperoglou. 'Aggregate Dataset on Descriptive Representation in the Hellenic Parliament (2019-2023)'. Aristotle University of Thessaloniki, 30 September 2024. <https://doi.org/10.5281/zenodo.13861613>.

Höhmänn, Daniel, and Ulrich Sieberer. "Parliamentary questions as a control mechanism in coalition governments." *West European Politics* 43, no. 1 (2020): 225-249.

Jacobs, Laura, Caroline Close, and Jean-Benoit Pilet. "The angry voter? The role of emotions in voting for the radical left and right at the 2019 Belgian elections." *International Political Science Review* (2024): 01925121231224524.

Martin, Shane, and Olivier Rozenberg, eds. *The roles and function of parliamentary questions*. Routledge, 2014.

Morales, Laura, Nayla Escribano, Alexandros-Christos Gotinakos, Matilde Ceron, Zoe Lefkofridi, Eftichia Teperoglou, and Nadine Zwiener-Collins. 'Data Structure, Codebook and Dictionaries of the Substantive Representation Datasets for WP4 of the Acteu Project'. Zenodo, 30 September 2024. <https://doi.org/10.5281/zenodo.13850648>.

Morales, Laura, and Carles Pamies. 'Aggregate Dataset on Descriptive Representation in the French Parliament (2012-2017)'. CSIC, 30 September 2024a. <https://doi.org/10.5281/zenodo.13862044>

Morales, Laura, and Carles Pamies. 'Aggregate Dataset on Descriptive Representation in the Italian Parliament (2013-2018)'. CSIC, 30 September 2024b. <https://doi.org/10.5281/zenodo.13862115>



Morales, Laura, and Carles Pamies. 'Aggregate Dataset on Descriptive Representation in the Spanish Parliament (2016-2023)'. CSIC, 30 September 2024c.

<https://doi.org/10.5281/zenodo.13862160>

Morales, Laura, and Carles Pamies. 'Aggregate Dataset on Descriptive Representation in the British Parliament (2015-2019)'. Zenodo, 30 September 2024d.

<https://doi.org/10.5281/zenodo.13862214>

Norris, Pippa. "Literature review and research paper on measuring trust." TrueDem project, (2023).

Palacios Brihuega, Irene. *Los españoles y la calidad de la democracia*. Vol. 74. CIS-Centro de Investigaciones Sociológicas, 2016.

Pitkin, Hanna F. *The concept of representation*. Univ of California Press, 1967.

Randour, Francois, Olivier Rozenberg, Jérémy Dodeigne, and Ferdinand Teuber.

"Understanding MEPs (varying) use of written parliamentary questions: domestic political culture and MEPs career patterns." (2022).

Rozenberg, Olivier, and Shane Martin. "Questioning parliamentary questions." In *The Impact of Legislatures*, pp. 446-456. Routledge, 2020.

Torcal, Mariano, and Pablo Christmann. "Responsiveness, performance and corruption: Reasons for the decline of political trust." *Frontiers in Political Science* 3 (2021): 676672.

Valgarðsson, Viktor, Will Jennings, Gerry Stoker, Hannah Bunting, Daniel Devine, Lawrence McKay, and Andrew Klassen. "A crisis of political trust? Global trends in institutional trust from 1958 to 2019." *British Journal of Political Science* (2024).

Van der Meer, Tom. W. G., and Patrick F. A. Van Erkel. "Moving beyond the political trust crisis debate: Residual analyses to understand trends in political trust." *European Journal of Political Research* 63, no. 3 (2024): 1240-1257.



Appendix – Links to codebooks, templates and other technical documentation of the datasets

This section provides the data citations and links for all the relevant documents used to generate the datasets on descriptive and substantive representation for WP4 of the ActEU project.

A.1 Codebooks and templates used for the descriptive representation datasets

Codebook of the Pathways project: Morales, Laura; Vintila, Daniela; Geese, Lucas; Mügge, Liza; van der Pas, Daphne; van de Wardt, Marc, 2017, "Codebook and data collection guidelines of work package 1 on descriptive political representation in national parliaments of the project Pathways to Power", <https://doi.org/10.7910/DVN/MRKJJF>, Harvard Dataverse, V1.

Codebook of the InclusiveParl project for LGBTI+ variables: Vincent-Mory, Claire; Morales, Laura, 2024, "Codebook and Data Collection Guidelines of the Project InclusiveParl on the Descriptive Representation of LGBTI+ Individuals", <https://doi.org/10.7910/DVN/HBKBMQ>, Harvard Dataverse, V1.

Vintila, Daniela, Laura Morales, and Claire Vincent-Mory. 'Excel Data Collection Template on Descriptive Political Representation in National Parliaments of the Projects Pathways to Power and Inclusiveparl Adapted for the Acteu Project'. Sciences Po, 30 September 2024. <https://doi.org/10.5281/zenodo.13859869>

Morales, Laura. 'Excel Template for the Aggregate Database on Descriptive Representation of the Acteu Project'. CSIC, 30 September 2024. <https://doi.org/10.5281/zenodo.13861463>

A.2 Codebooks and templates used for the substantive representation datasets

Morales, Laura, Nayla Escribano, Alexandros-Christos Gotinacos, Matilde Ceron, Zoe Lefkofridi, Eftichia Teperoglou, and Nadine Zwiener-Collins. 'Data Structure, Codebook and Dictionaries of the Substantive Representation Datasets for WP4 of the ActEU Project'. CSIC, 30 September 2024. <https://doi.org/10.5281/zenodo.13850648>

Escribano, Nayla, and Laura Morales. 'Aggregate Dataset on Substantive Representation in the Austrian, British, French, Italian and Spanish Parliaments (2013-2024)'. CSIC, 30 September 2024. <https://doi.org/10.5281/zenodo.13862300>



A.3 List of keywords in English used for the substantive representation datasets

Table 11 Description of the variables included in the descriptive representation dataset (Laura Morales and Nayla Escribano, CSIC)

WOMEN'S ISSUES DICTIONARY	MEN'S ISSUES DICTIONARY	TRANS ISSUES DICTIONARY	NON BINARY ISSUES DICTIONARY	LGBTI+ ISSUES DICTIONARY	YOUTH ISSUES DICTIONARY	IMMIGRATION ISSUES DICTIONARY	MINORITY ISSUES DICTIONARY	CLIMATE CHANGE ISSUES DICTIONARY
English	English	English	English	English	English	English	English	English
Woman	Man	Trans person	Non-binary	LGB / LGBT / LGBTI / LGBTQ / LGBTQ+ / LGBTQIA+ / ...	Youth	Immigration	Ethnic minority	IPCC
Girl	Boy	Trans woman	Androgyny	Come out	Young	Immigrant	Islamophobia	greenhouse gas
Feminine	Masculine	Trans man	Androgynous	Outing	Minor	Migration	Islamophobic	anthropogenic emissions
Femininity	Masculinity	Transgender	Two-spirit	Rainbow flag	Adolescent	Migrant	Antisemitism	gas emission
Childcare	Manhood	Transsexual	Gender-fluid	Pride month	Adolescence	Refugee	Antisemitic	emit gases
Parenting	Father	Gender transition	Fluid sexuality	Pride parade	Teenager	Asylum	Racism	weather
Mother	Paternity	Gender recognition	Bigender	Marginalized orientation	Girl	Xenophobia	Racist	climate
Maternity	Paternal	Gender self-identification	Gender non-conforming	Gender identity	Boy	Xenophobic	Multiculturalism	climatic



Maternal	Parental leave	Gender expression	Gender questioning	Sexual identity	Child	Family reunification	Multicultural	ecosystem
Parental leave	Violence against men	Sex reassignment		Sexual orientation	Kid	Reunite families	Diaspora	environment
Abortion	Discrimination against men	Gender reassignment		Sexual preference	Childhood	Settlement		environmental
Abort	Single father	Hormone replacement		same-sex	Infancy	Settle down		biodiversity
Pregnancy	Misandry	Transphobia		Homosexual	Infantile	Remittances		temperature
Pregnant		Transphobic		Homosexuality	Age of consent			warming
Violence against women		Transmisogyny		Heterosexual	Gen-Z			heatwave
Male violence		Transmisogynist		Heterosexuality	Juvenile			drought
Sexual harassment		TERF		Lesbian	New generation			wildfire
Gender violence		Misgendering		Lesbianism				flood
Gender-based violence		Misgender		Bisexual				glacier
Domestic violence		Gender-affirming		Bisexuality				glacial
Family violence		Gender-neutral		Pansexual				food security



Domestic abuse		Gender neutrality		Pansexuality				food insecurity
Spousal abuse		Inclusive language		Omnisexual				water security
Rape				Omnisexuality				water insecurity
Sexual abuse				Asexual				water availability
Stalking				Asexuality				available water
Stalk				Heteroflexible				ocean acidification
Street harassment				Heteroflexibility				coastal damage
Femicide				Homoflexible				afforestation
Female genital mutilation				Homoflexibility				deforestation
Feminism				Queer				acidic rain
Feminist				Genderqueer				CO2
Sexism				Homophobic				carbon dioxide
Sexist				Homophobia				habitat destruction



Misogyny				Lesbophobic				emission target
Misogynist				Lesbophobia				carbon neutral
Patriarchy				Biphobic				carbon neutrality
Patriarchal				Biphobia				water scarcity
Gender equality				Queerphobic				water stress
Gender balance				Queerphobia				carbon footprint
Gender parity								renewable energy
Gender pay gap								natural disaster
Gender gap								sustainability
Gender disparity								sustainable
Gender inequality								ocean level
Gender imbalance								sea level
Gender discrimination								ice melting



Sex discrimination								
Glass ceiling								
Reproductive rights								
Reproductive health								
Menstrual								
Menstruation								
Postpartum								
Childbirth								
Give birth								
Contraception								
Contraceptive								
Birth control								



About ActEU

How can we conceptualize and empirically measure political trust and legitimacy beyond the usual survey question “How much trust do you have in the parliament?”? Does the multi-level nature of European representative democracies require an identical level of citizen support at the regional, national and EU levels? How does social polarization on key policy issues of our times –immigration, climate change, and gender inequality– challenge the political trust in, and legitimacy of, democratic political systems? And what can policymakers and civil society do to master these challenges? ActEU aims at finding answers to these questions pursuing two overarching goals: In phase 1, we map and investigate persistent problems of declining trust, legitimacy and representation in Europe with a particular attention to the polarization of societies and the EU’s multi-level structures. Providing an innovative conceptual framework on political attitudes, behavior and representation across Europe, we establish an original empirical infrastructure based on an innovative combination of methods and newly collected quantitative and qualitative empirical data (focus groups, experimental surveys, web scraping). In phase 2, these results will flow directly into the creation of a toolbox of remedial actions to enhance political trust in and legitimacy of European representative democracies. In cooperation with a newly created Civil Society Network, Youth Democracy Labs across 13 European cities and in exchange with political cartoonists “Cartooning for democracy”, we will develop context-sensitive solutions for all polity levels and some of the most polarizing policy areas, and craft tailor-made toolkits for both policymakers and civil society and the educational sector. Finally, we deploy a differentiated dissemination strategy to maximize ActEU’s scientific, policy and societal impact in activating European citizens’ trust and working towards a new era of representative democracy.

www.acteu.org

ActEU Consortium



License This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.