

**Comprehensive
European Parliament
Electoral Data
(COMEPELDA)**

Codebook

*Codebook version: 1.0
(September 22, 2021)*

Authors: Thomas Däubler and Mihail Chiru

The most recent version of the data should be available from:
<https://doi.org/10.7910/DVN/GNRMTO>

CITATION

When using the data please cite:

Däubler, Thomas, Chiru, Mihail and Hermansen, Silje Synnøve Lyder 2021. “Introducing COMEPELDA: Comprehensive European Parliament Electoral Data covering rules, parties and candidates”. European Union Politics. <https://doi.org/10.1177%2F14651165211053439>

The data includes and integrates larger amounts of information from the following sources. We recommend that end users also cite these sources when using the respective variables:

- National-level election results:
 - Döring, Holger and Philip Manow (2020). “Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in modern democracies”. Development version. Available at parlgov.org.
- Information on MEPs’ parliamentary spells:
 - Høyland, Bjørn, Sircar, Indraneel, and Simon Hix. (2009). “An Automated Database of the European Parliament”. *European Union Politics*, 10 (1): 143-152.
- Information on electoral rules for 2004-2014:
 - Däubler, Thomas, and Simon Hix. 2018. ”Ballot Structure, List Flexibility and Policy Representation.” *Journal of European Public Policy* 25 (12): 1798-1816.

ACKNOWLEDGEMENTS

The data collection endeavor would not have been possible without the diligent research assistance by Ertan Bat. We are grateful to him. We also would like to thank all the people who helped with country-specific questions and other queries. These include Patrick Dumont, Kostas Gemenis, George Hajipavli, Bjørn Høyland, Raimondas Ibenskas, Tom Louwerse, Julien Navarro and Tevfik Murat Yildirim. We apologize to anybody we have forgotten to mention.

CONTACT

If you spot any errors in the data or the codebook, or if you have any questions that are not answered here, please send an email (using COMEPELDA in the subject) to thomas.daubler@ucd.ie.

Contents

1	Overview	1
2	Lists of variables in the three datasets	2
3	Detailed notes on the variables	7
3.1	Aggregate data (AD)	7
3.2	MEP data (MD)	12
3.3	Candidate data (CD)	16
4	Combining the three datasets (R examples)	18
5	Parties without ParlGov ID	21
6	Main sources of electoral data	22

List of Tables

1	Aggregate data (AD)	3
2	Aggregate data (AD) - continued	4
3	MEP data (MD)	5
4	Candidate data (CD)	6
5	Parties without ParlGov ID	21
6	Main data sources of election results (links live as of summer 2021)	22
7	Main data sources of election results (continued)	23

1. Overview

- Coverage:
 - 'Regular' EP elections: 1999, 2004, 2009, 2014
 - 'By-elections' related to EU accession: 2007 (Bulgaria, Romania), 2013 (Croatia)
- There are three separate datasets:
 - Aggregate Dataset (AD)
 - MEP Dataset (MD)
 - Candidate Dataset (CD)
- File format: comma-separated values (take this literally), using UTF-8 text encoding; variable names in first line; dots (.) for missing values; and double quotation marks for quotes within string variables
 - To read into R you can use, e.g.:

```
da <- read.csv(file="aggregate_v0.99.csv",
               header=TRUE, encoding = "UTF-8", na.strings=".")
```
 - To read into Stata you can use (this may require adding 'clear' as an argument if you need to and want to overwrite the data in memory), e.g.:

```
import delim aggregate_v0.99.csv,
varnam(1) encoding("UTF-8") case(preserve)
```
 - To keep the data easy to read with various software packages, we do not distinguish between different types of missing values in the data files themselves. Further information on the non-applicability of certain variables is given in [Section 3](#) below.
- Each dataset has a numeric (artificial) key to identify its rows. In some cases alternate keys can be constructed based on the combination of several natural variables. See below for further information. [Section 4](#) below provides some examples using R.

2. Lists of variables in the three datasets

Preliminaries:

- Identifying variables begin with upper case **ID**.
- Lower-case opening letters in variable names stand for: **c=country, p=party, d=district**.
- All electoral data is election year-specific, but for simplicity this is not reflected in variable names.
- Type of variable: **bin=binary, char=character, num=numeric**

Table 1. Aggregate data (AD)

Variable	Description	Type
Identifiers and affiliation		
IDAD	numeric row ID for Aggregate Data	num
cName	Country name	char
EIYear	Election year	num
IDpty	Party ID (ParlGov)	num
pName	Party name	char
dName	District name	char
dNameEng	District name English	char
Votes		
cValid	Valid votes (country)	num
dValid	Valid votes (district)	num
pcAbs	Number of votes (party-country)	num
pcShare	Vote share (party-country)	num
pdAbs	Number of votes (party-district)	num
pdShare	Vote share (party-district)	num
Candidates and Seats		
pdNofCandSel	Number of candidates (selection, party-district)	num
pdNofCandEl	Number of candidates (election, party-district)	num
cSeats	Seats in EP (country)	num
pdFlagNofCand01	Indicator for potential errors in these variables	bin
dMag	District magnitude	num
pcSeats	Seats (party-country)	num
pdSeats	Seats (party-district)	num
Rules		
pDistricted01	Subnational district use	bin
cFormula	Electoral formula (between-party)	char
cBetwPtyThresh	Legal threshold	char
cAppar01	Possibility of apparentement	bin
cMultiTier01	Multi-tier seat allocaton	bin
cFormulaBetwDist	Electoral formula (between-districts)	char
cBallotType	Ballot Structure	char
cPanach01	Use of panachage	bin
cRankedList01	Party-determined pre-electoral ranking	bin
cPrefVotOblig01	Preference voting obligatory	bin
cPrefVotN	Number of preference votes	num
cPrefThrFormal	Preference vote threshold rule	char

Table 2. Aggregate data (AD) - continued

Variable	Description	Type
Preference votes		
pdPrefThrEmp	Preference vote threshold, empirical, absolute votes	num
pdPrefVotSum	Preference vote sum	num
pdPrefVotLoser	Preference vote max of non-elected candidates	num
pdListFlex	List flexibility	char
pdPat	Post-electoral allocation type	char
Miscellaneous		
IndepCand01	Independent candidate	bin
cAddLisbon	Additional seats Lisbon treaty (country)	num
pAddLisbon	Additional seats Lisbon treaty (party)	num
pAlliance01	Party alliance	bin
IDauxAlliance	Alliance ID (ParlGov)	num
pApparParty1	Party ID of apparentement partner 1	num
pApparParty2	Party ID of apparentement partner 2	num
IDauxElec	Election ID (ParlGov)	num
IDauxIndep	Party ID (ParlGov) independent candidates	num

Table 3. MEP data (MD)

Variable	Description	Type
Identifiers		
IDMD	numeric row ID for MEP data	num
IDmep	MEP ID (assigned by EP)	num
IDauxWik	Wikidata person ID	num
EP	EP session	num
NofEpis	Number of mandate episodes	num
Episode	Episode index	num
LastName		char
FirstName		char
Episode information		
Start	Start date of mandate episode (YYYY-MM-DD)	char
End	End date of mandate episode (YYYY-MM-DD)	char
EpisType	Type of mandate episode	char
cName	Country name (electoral)	char
MemberState	Delegation	char
Previous election		
EIYearLag	Election year	num
Run01Lag	Running	bin
MultCdy01Lag	Multiple candidacies	bin
IDCDfocLag	ID of focal candidacy	num
IDADfocLag	ID of Aggregate Data entry for focal candidacy	num
IDptyLag	Party ID	num
dNameLag	District name	num
Elected01Lag	Initially elected	bin
Elected01MLag	Initially elected, (not) considering multiple election instances	bin
Following election		
EIYearLead	Election year	num
Run01Lead	Running	bin
MultCdy01Lead	Multiple candidacies	bin
IDCDfocLead	ID of focal candidacy	num
IDADfocLead	ID of Aggregate Data entry for focal candidacy	num
IDptyLead	Party ID	num
dNameLead	District name	char
ElectedAny01Lead	Initially elected in any district	bin
ElectedDis01MLead	Initially elected in represented district (not) considering multiple elections	bin
Miscellaneous		
InEP4 [5/6/7/8]	MEP in EP4 [5/6/7/8]	bin

Table 4. Candidate data (CD)

Variable	Description	Type
Identifiers and affiliation		
IDCD	numeric row ID for candidate data	num
IDmepAdjSess	MEP ID (if MEP in an adjacent, i.e. the outgoing or the next, session)	num
IDMDnextall	Row ID MEP data (if MEP in next session)	num
IDMDnextrep	Row ID MEP data (if this candidacy results in mandate in next session)	num
IDMDoutall	Row ID MEP data (if MEP in outgoing session)	num
IDMDoutrep	Row ID MEP data (if MEP in outgoing session for this district)	num
IDAD	Row ID Aggregate data	num
EIYear	Election year	num
IDpty	Party ID (ParlGov)	num
cName	Country name	char
dName	District name	char
dNameEng	District name English	char
SuccList01	Successor list	bin
Candidate information		
Name	Candidate name	char
MultCdy01	Multiple candidacies	bin
ListRankSel	List rank at selection	num
ListRankEl	List rank at election	num
FlagListRank01	Indicator for errors in list rank variables	bin
PrefVot	Preference votes	num
PrefVotLatvia	Positive minus negative votes	num
Elected01	Initially elected	bin
Elected01M	Initially elected, (not) considering multiple election instances	bin
OrderElec	Order of election	num

3. Detailed notes on the variables

3.1 Aggregate data (AD)

Identifiers and affiliation

- IDAD
 - Numeric row ID (without substantive meaning, serving as the primary key).
 - It implies a unique combination of *IDpty*, *EIYear* and *dName*.
- cName
 - Country name, which refers to the 'country' in the sense of an electoral context.
 - Therefore, we distinguish Great Britain and Northern Ireland.
- EIYear
 - Election year.
 - The data provides some variables for the MEPs appointed in Bulgaria and Romania from January 2007 until the by-elections of the same year. These observations have *EIYear*==2006.
- IDpty
 - Party ID from ParlGov.
 - We followed ParlGov in dealing with alliances (see <http://www.parlgov.org/documentation/codebook/#party>), i.e. *IDpty* refers to the ID that ParlGov uses in listing the results from the EP election of interest. See also variables *pAlliance01* and *IDauxAlliance*.
 - For parties that do not have a ParlGov ID, we created IDs starting with -9901 and counting down. See Section 5 for a list.
 - Since ParlGov does not always use person-specific IDs for independent (i.e. running on their own) candidates, we created IDs by concatenating -88 and their *IDmep* (this works since we do not cover any independent candidates who were never MEPs.)
- pName
 - Party name.
 - This is the local language name, typically from ParlGov, referring to *IDpty*. Note that this need not be the name under which the party registered for the elections or can be found in the official results. Also see the remarks on alliances in the notes for *IDpty*.
- dName
 - District name in local language.
 - It includes two-letter Eurostat country codes to improve clarity in some cases. Note that there can be boundary changes over time, which may or may not be reflected in changes of the district names.
- dNameEng
 - District name in English.
 - It includes two-letter Eurostat country codes to improve clarity in some cases. Note that there can be boundary changes over time, which may or may not be reflected in changes of the district names.

Votes

- *cValid*
 - Absolute number of valid votes in country (i.e. *cName*).
 - Typically from ParlGov. For Italy 1999 and France 2004 we include different figures corresponding to the sum of *dValid*.
- *dValid*
 - Absolute number of valid votes in district
- *pcAbs*
 - Absolute number of votes for party in country (i.e. *cName*).
 - Typically from ParlGov.
- *pcShare*
 - Proportion of votes for party in country (i.e. *cName*)
 - Typically from ParlGov.
 - For Luxembourg, with the panachage system, we use the sum of preference votes across parties in the denominator. Since we re-calculated this, there are minor differences to the vote share figures ParlGov reports for Luxembourg.
- *pdAbs*
 - Absolute number of votes for party in district
- *pdShare*
 - Proportion of votes for party in district.
 - Also see the notes for *pcShare*.

Candidates and Seats

- *pdNofCandSel*
 - Number of candidates for party in district, at selection stage.
 - This differs (in rare cases) from *pdNofCandEl* if we are aware of candidates who withdrew, were denied registration or passed away before the elections.
- *pdNofCandEl*
 - Number of candidates for party in district, at electoral stage.
 - This differs (in rare cases) from *pdNofCandSel* if we are aware of candidates who withdrew, were denied registration or passed away before the elections.
- *pdFlagNofCand01*
 - Indicates potential errors in the *pdNofCand** variables.
 - This variable equals one if there is reason to believe that the raw data at hand may not include all the candidates. End users may want to exclude such parties.
- *cSeats*

- Number of seats in EP for country (i.e. *cName*)
- dMag
 - District magnitude.
 - For Italy and Poland, this is the number of seats allocated in the district ex post, i.e. based on the election results (it is not fixed ex ante).
 - The German case is not straightforward to code, since parties can choose whether to use nation-wide or district-level lists. Among German parties we cover, only the CDU and CSU use (connected) district-level lists. In the data we code *dMag* as *pdSeats* (or as 1 if *pdSeats*==0) for CDU and CSU, and as *cSeats* minus the country-level sum (*pcSeats*) of CDU/CSU for the other parties. We do not claim that this is the best way of solving the issue or that it fits every research question. It is simply a way of providing information potentially useful to some end users, rather than missing values.
- pcSeats
 - Number of seats for party in country (i.e. *cName*)
- pdSeats
 - Number of seats for party in district

Rules

- pDistricted01
 - Indicates whether subnational districts are used.
 - This is 0 for Northern Ireland, since it has its own *cName*). The variable prefix reflects the fact that this can vary at the party level for Germany, since parties can decide whether to submit nation-wide or district-level lists.
- cFormula
 - Description of electoral formula for between-party seat allocation
- cBetwPtyThresh
 - Description of legal threshold for between-party seat allocation.
 - Note that the legal threshold may be ineffective if it is lower than the natural threshold.
- cAppar01
 - Indicates the possibility of apparentement in country
- cMultiTier01
 - Indicates the presence of multi-tier seat allocation
- cFormulaBetwDist
 - Description of electoral formula for between-district seat allocation
- cBallotType
 - Ballot Structure.
 - Categories: closed lists, flexible lists, open lists , and STV. See Däubler and Hix (2018).

- cPanach01
 - Indicates if panachage (preference voting for candidates from different lists is allowed). This is not coded for STV (which is a non-list system), which allows to rank order any candidates, or for closed list systems.
 - Does Not Apply (DNA) if BallotType == closed or STV
- cRankedList01
 - Indicates if the party ranks candidates on the list before the elections.
 - Compared to Däubler and Hix (2018), there are two corrections: Denmark and Luxembourg are coded as 1.
 - DNA if BallotType == STV
- cPrefVotOblig01
 - Indicates whether preference voting is obligatory
 - DNA if BallotType == closed
- cPrefVotN
 - Maximal number of preference votes that are admissible
 - DNA if BallotType == closed
- cPrefThrFormal
 - Description of the preference vote threshold rule. See Däubler and Hix (2018).
 - DNA if BallotType != flexible

Preference votes

- pdPrefThrEmp
 - Ex post preference vote threshold in absolute votes for party in district
 - It is not rounded to the next largest integer, so candidates must exceed the threshold.
 - DNA if BallotType != flexible
- pdPrefVotSum
 - Sum of preference votes for party in district
 - DNA if BallotType == closed
- pdPrefVotLoser
 - Preference vote result of best non-elected candidate for party in district.
 - For STV systems, this is based on first preferences. This is not available for some parties for the Swedish 1999 elections (compare the notes on *PrefVot* in Candidate Dataset).
 - DNA if BallotType == closed or pdSeats == 0 or pdNofCandEl == pdSeats
- pdListFlex
 - List flexibility
 - Weakly vs strongly flexible, ex post, for party in district, see Däubler and Hix (2018).

- DNA if BallotType != flexible or pdSeats == 0
- pdPat
 - Post-electoral allocation type, ex post, for party in district: list rank (LR)-based, personal vote (PV)-based or mixed (MX), see Däubler, Thomas. 2020. 'Do More Flexible Lists Increase the Take-up of Preference Voting?'. Electoral Studies 68.
 - DNA if BallotType == STV or pdSeats == 0
- Miscellaneous**
- IndepCand01
 - Indicates if the entry refers to an independent candidate running on his/her own.
 - This is zero for teams of candidates who call themselves independents but whose votes pool (like the Czech 'Nezávislí').
- cAddLisbon
 - Number of additional seats the country received in the context of the Lisbon Treaty-related enlargement of the EP
- pAddLisbon
 - Number of additional seats the party received in the context of the Lisbon treaty-related enlargement of the EP
- pAlliance01
 - Indicates if the entry refers to an alliance of parties (ParlGov)
- IDauxAlliance
 - Alliance ID from ParlGov.
 - Can be used with the ParlGov table with title election result to look up the component parties. This can be missing even if pAlliance01==1 if ParlGov does not cover the case in the specific election. Note that the ParlGov alliance numbering/ID is different from the ParlGov party numbering/ID.
 - DNA if pAlliance01 == 0
- pApparParty1
 - Party ID of partner party 1 if party uses list connections/apparentement. Small parties without ParlGov ID are not listed. Note that we included those connections that we are aware of, but we did not invest much time in checking for completeness.
 - DNA if cAppar01 == 0
- pApparParty2
 - Party ID of partner party 2 if party uses list connections/apparentement. Small parties without ParlGov ID are not listed. Note that we included those connections that we are aware of, but we did not invest much time in checking for completeness.
 - DNA if cAppar01 == 0
- IDauxElec
 - Election ID from ParlGov.

- Here, Great Britain and Northern Ireland both receive the UK election ID from ParlGov.
- IDauxIndep
 - This is the Party ID that ParlGov uses for the independent candidate in the given election (compare the remarks on the *IDpty* variable).
 - DNA if *IndepCand01*==0

Note: For the MEPs appointed in Romania and Bulgaria for January 2007 (*EIYear*==2006), most of the variables above are not applicable. For simplicity, this is not indicated above.

3.2 MEP data (MD)

Identifiers

- IDMD
 - Numeric row ID (without substantive meaning, serving as the primary key)
 - It implies a unique combination of *IDmep*, *EP* and *Episode*.
- IDmep
 - Person ID for the MEP as assigned by EP
- IDauxWik
 - Wikidata person ID.
 - This is missing only for a handful of politicians who apparently lack such an ID.
 - The WikiData page of the MEP can be accessed by pasting the ID at the end of: <https://www.wikidata.org/wiki/>
- EP
 - EP session
 - (5 for 1999-2004, 6 for 2004-2009, 7 for 2009-2014, 8 for 2014-2019)
- NofEpis
 - Number of mandate episodes for the MEP in that session
 - Compare notes on *EpisType* below.
- Episode
 - Index for the mandate episodes of an MEP in that session
- LastName
 - MEP family name
- FirstName
 - MEP forename

Episode information

- Start

- Start date of mandate episode
- End
 - End date of mandate episode
- EpisType
 - Mandate type of episode.
 - Categories: 'init elec'=initially elected; 'repl begin'=replacement from beginning of session; 'repl sess'=replacement who started during the session; 'app EU enl'=appointed in the context of EU enlargement; 'app Lisb'=appointed in the context of the Lisbon Treaty enlargement of the EP; 'observ'=observer.
 - The rare cases of initially elected candidates who initially do not take up the seat but become replacements later are coded as 'repl sess'.
- cName
 - Country name, which refers to the 'country' in the sense of an electoral context.
 - Therefore, we distinguish Great Britain and Northern Ireland.
- MemberState
 - Member state.
 - Equals United Kingdom if *cName* refers to Great Britain or Northern Ireland.

Previous election

As the suffix 'Lag' indicates, all variables in this section refer to the previous election

- ElYearLag
 - Election year.
 - This is missing for EpisType=='observ', and for EpisType=='app EU enl' (except for the Bulgarian and Romanian cases with Start=='2007-01-01', who have entries in the Aggregate Data for that appointment process).
 - DNA if no previous EP election took place in this country
- Run01Lag
 - Indicates if the MEP was a candidate.
 - This is missing if the country did not hold an EP election between EU accession and the *Start* date of the episode (i.e. also for the Bulgaria and Romania 2006 exceptions discussed in the Notes for *ElYearLag*).
 - While rare, it is possible that replacement or appointed MEPs obtain a mandate although an EP election was held before and they did not run in it (typically, these are replacement MEPs in STV systems or Lisbon Treaty appointments).
 - DNA if no previous EP election took place in this country
- MultCdy01Lag
 - Indicates if the MEP had multiple candidacies (in several districts or on both the main and the successor list)
 - DNA if Run01Lag == 0

- IDCDfocLag
 - ID of focal candidacy in Candidate Dataset. If the MEP had multiple candidacies (MultCdy01Lag==1), this refers to the district the MEP eventually represented and the list type (main vs successor) the mandate was based on.
 - DNA if Run01Lag == 0
- IDADfocLag
 - ID of entry in Aggregate Dataset that corresponds to the focal candidacy
 - DNA if Run01Lag == 0
- IDptyLag
 - Party ID from ParlGov.
 - See Notes for Aggregate Data for further details.
 - DNA if Run01Lag == 0
- dNameLag
 - Name of the district the MEP represents, in local language.
 - This includes two-letter Eurostat country codes to improve clarity in some cases.
 - DNA if Run01Lag == 0
- Elected01Lag
 - Indicates if MEP was initially elected
 - DNA if Run01Lag == 0
- Elected01MLag
 - Indicates if MEP was initially elected, but codes MEPs succeeding for those candidates who were elected in more than one district as 0 (such cases have EpisType=='init elec' but can be identified with this variable).
 - DNA if Run01Lag == 0

Following election

As the suffix 'Lead' indicates, all variables in this section refer to the following election

- ElYearLead
 - Election year.
 - DNA if EpisType == 'observ', which is not followed by an election (but by an appointment round. Exception: Bulgarian and Romanian cases with End=='2006-12-31', who have entries in the Aggregate Data for the subsequent appointment process)
- Run01Lead
 - Indicates if the MEP was a candidate, i.e. appears in the Candidate Data in the next election.
 - This is not yet coded for ElYearLead==2019.
 - While we have invested considerable time in checking this information, we may code some cases incorrectly (e.g., when MEPs end up running for parties receiving very few votes that we fail to include). Note that we do not consider reasons for not running again, e.g., any MEPs who passed away before the next elections would still receive a value of 0.

- DNA if EpisType == 'observ', which is not followed by an election
- MultCdy01Lead
 - Indicates if the MEP had multiple candidacies (in several districts or on both main and successor list)
 - DNA if Run01Lead == 0
- IDCDFocLead
 - ID of focal candidacy in Candidate Dataset.
 - If the MEP had multiple candidacies (MultCdy01Lead==1):
 - * on the main and successor list, this refers to the main list;
 - * in several districts, this refers to the district the MEP represented in the outgoing term.
 - There is only one case of an MEP (IDMD==482) running in more than one district but not the outgoing one (for this case we coded the focal district based on the preference vote result).
 - DNA if Run01Lead == 0
- IDADfocLead
 - ID of entry in Aggregate Dataset that corresponds to focal candidacy
 - DNA if Run01Lead == 0
- IDptyLead
 - Party ID from ParlGov.
 - See Notes for Aggregate Data for further details.
 - DNA if Run01Lead == 0
- dNameLead
 - Name of the district for the focal candidacy, in local language
 - DNA if Run01Lead == 0
- ElectedAny01Lead
 - Indicates if MEP is initially elected in any district
 - DNA if Run01Lead == 0
- ElectedDis01MLead
 - Indicates if MEP is initially elected in the district represented in the outgoing EP session, allowing for multiple election. This codes MEPs succeeding for those candidates who were elected in more than one district as 0.
 - DNA if Run01Lead == 0

Miscellaneous

- InEP4 [5/6/7/8]
 - Indicates if MEP had any mandate episode in EP 4 [5/6/7/8]

3.3 Candidate data (CD)

Identifiers and affiliation

- IDCD
 - Numeric row ID (without substantive meaning, serving as the primary key).
 - It implies a unique combination of *EIYear*, *IDpty*, *dName*, *SuccList0* and *Name*.
- IDmepAdjSess
 - This is the MEP ID (as assigned by the EP itself).
 - Currently, this variable has non-missing values only if the candidate had a mandate episode in an adjacent EP session, i.e. the outgoing or following legislative term.
 - DNA if not an MEP in the EP session before or after the election
- IDMDnextall
 - This refers to the Row ID in the MEP dataset (MD).
 - This variable has non-missing values only if the candidate had a mandate episode in the next EP session. In case of multiple candidacies, it has values in all of them.
 - DNA if not an MEP in the next session
- IDMDnextrep
 - This refers to the Row ID in the MEP dataset (MD).
 - This variable has non-missing values only if the candidate had a mandate episode in the next EP session and only for the district and list type (main vs successor) the following mandate was based on.
 - DNA if not an MEP in the next session for this district
- IDMDoutall
 - This refers to the Row ID in the MEP dataset (MD).
 - This variable has non-missing values only if the candidate had a mandate episode in the outgoing EP session (note that this need not mean that the MEP was in office at the end of the session). In case of multiple candidacies, this holds values for all of them.
 - DNA if not an MEP in the previous session
- IDMDoutrep
 - This refers to the Row ID in the MEP dataset (MD).
 - This variable has non-missing values only if the candidate had a mandate episode in the outgoing EP session (note that this need not mean that the MEP was in office at the end of the session). In case of multiple candidacies in the current election, there is a value only for the district represented in the outgoing EP but (where applicable) for both the main and successor list.
 - When dealing with the Irish district boundary changes, we considered MEPs as outgoing as long as the new district included non-negligible parts of the area of the previous one. For France 2004 (nationwide district in 1999) and Poland 2004 (first EP election) all cases with entries in *IDMDoutall* have values also in this variable.
 - DNA if not an MEP in the previous term for this district

- IDAD
 - This refers to the Row ID in the Aggregate Dataset (AD).
 - It implies a unique combination of *IDpty*, *EIYear* and *dName*.
- EIYear
 - Election year
- IDpty
 - Party ID from ParlGov.
 - See Notes for Aggregate Data for further details.
- cName
 - Country name.
 - See Notes for Aggregate Data for further details.
- dName
 - District name.
 - See Notes for Aggregate Data for further details.
- dNameEng
 - District name English.
 - See Notes for Aggregate Data for further details.
- SuccList01
 - Indicates if the list was a separate successor list (from which candidates replacing MEPs are recruited) rather than a main list.
 - For the 2004 elections in Portugal, the successor lists are not at hand.
 - For IDpty==1273 in 1999 (IDAD==1656) the successor list cases at hand seem few.

Candidate information

- Name
 - Candidate name
 - Currently, the format and content in this variable varies across countries and elections; it is planned that future versions of the dataset will present this information in a more standardized manner.
- MultCdy01
 - Indicates if the MEP had multiple candidacies (in several districts or on both main and successor list)
- ListRankSel
 - List rank at selection stage.
 - This differs (in rare cases) from *ListRankEl* if we are aware of candidates who withdrew, were denied registration or passed away before the elections.

- DNA if `cRankedList01[cName] == 0`
- ListRankEl
 - List rank at election stage.
 - This will be missing for candidates who are known to have withdrawn, been denied registration or passed away before the elections.
 - DNA if `cRankedList01[cName] == 0`
- FlagListRank01
 - Indicates that this is a list for which list rank information is missing for some less prominent candidates (mostly just one) but retained (as a slightly erroneous proxy) for the others.
 - DNA if `cRankedList01[cName] == 0`
- PrefVot
 - Preference votes for the candidates.
 - For STV systems, this holds first preferences. For Latvia, this holds the number of positive preferences. For the Sweden 1999 elections, the information was published only for candidates who were initially elected or cleared the threshold. Otherwise the information is missing only for a handful of cases.
 - DNA if `cBallotType[cName] == closed`
- PrefVotLatvia
 - The number of positive preferences minus the number of negative preferences for Latvian candidates.
 - DNA if `cName != Latvia`
- Elected01
 - Indicates if the candidacy resulted in (initial) election.
 - DNA if we know that the candidate withdrew, was denied registration or passed away before the election.
- Elected01M
 - Indicates if the candidacy resulted in election. This variant codes instances of candidates succeeding for those candidates who were elected in more than one district (who will have several entries of 1 in the given election) as 0.
 - DNA if we know that the candidate withdrew, was denied registration or passed away before the election.
- OrderElec
 - The within-party order of election, for those candidates initially elected.
 - DNA if `Elected01 == 0`

4. Combining the three datasets (R examples)

Beginning on the next page, this section presents some examples that illustrate how to merge/join the datasets. While the code is for R, we hope that users of other software may still find them useful ('by.x' specifies the key variables in the left dataset and 'by.y' in the right dataset; simple 'by' assumes the key variables have the same name in both datasets).

All examples assume the three datasets were read in as the objects: *da* (aggregate), *dc* (candidates), and *dm* (MEPs)

Example 1: Add info from Aggregate Data to Candidate Data

Here: absolute party votes (and number of candidates)

```
dana <- merge(dc,
             da[,c("IDAD", "pdAbs", "pdNofCandEl")],
             by = "IDAD",
             all.x=TRUE)
```

Share of all party voters supporting the candidate:

```
dana$pvsupport <- dana$PrefVot/dana$pdAbs
summary(dana$pvsupport)
```

Exclude instances when there is just one candidate:

```
summary(dana$pvsupport[dana$pdNofCandEl > 1])
```

Look at candidates with > 90% support:

```
View(dana[!is.na(dana$pvsupport) & dana$pvsupport > .9 & dana$pdNofCandEl > 1,
        c("cName", "EYear", "IDpty", "dName", "Name", "pdAbs", "PrefVot", "pvsupport")])
```

Example 2: Add info from Aggregate Data to MEP Data

Here: vote share of party in district in *previous* election

```
dana <- merge(dm,
             da[,c("IDAD", "pdShare")],
             by.x="IDADfocLag",
             by.y="IDAD",
             all.x=TRUE)
summary(dana$pdShare)
```

Here: vote share of party in district in *following* election

```
dana <- merge(dm,
             da[,c("IDAD", "pdShare")],
             by.x="IDADfocLead",
             by.y="IDAD",
             all.x=TRUE)
summary(dana$pdShare)
```

Example 3: Add info from Candidate Data to MEP Data

Here: list rank related to previous and following election

```
dana <- merge(dm,
             dc[,c("IDCD", "ListRankSel")],
             by.x="IDCDfocLag",
             by.y="IDCD",
             all.x=TRUE)
names(dana)[names(dana)=="ListRankSel"] <- "ListRankSelLag"

dana <- merge(dana,
             dc[,c("IDCD", "ListRankSel")],
             by.x="IDCDfocLead",
             by.y="IDCD",
```

```

      all.x=TRUE)
names(dana)[names(dana)=="ListRankSel"] <- "ListRankSelLead"

summary(dana[, c("ListRankSelLag", "ListRankSelLead")])

```

Basic (!) comparison of list ranks over time:

```
table(sign(dana$ListRankSelLead - dana$ListRankSelLag))
```

Note that this will be missing if parties don't rank lists, and the comparison over time applies only to MEPs who have run in both the previous and the following election. In addition, this simple analysis ignores that MEPs may switch parties, list types (main vs successor), or districts.

Example 4: Add info from MEP Data to Aggregate Data

Here: relative frequency of replacing MEPs in the party during the previous term

```

library(dplyr)
tmp <- summarize(group_by(dm, EP, IDADfocLead),
  Ninit = sum(EpisType %in% c("init elec", "repl begin")), # MEPs after election
  Nrepl = sum(EpisType == "repl sess")) # replacements

dana <- merge(da,
  tmp[tmp$Ninit > 0, c("IDADfocLead", "Ninit", "Nrepl")],
  by.x="IDAD",
  by.y="IDADfocLead",
  all.x=TRUE)

table(dana$Nrepl)

```

Distribution of replacement rates:

```
summary(dana$Nrepl/dana$Ninit)
```

Example 5: Add info from MEP Data to Candidate Data

Here: episode type related to outgoing term

```

dana <- merge(dc,
  dm[, c("IDMD", "EpisType")],
  by.x="IDMDoutall",
  by.y="IDMD",
  all.x=TRUE)
table(dana$EpisType, useNA = "ifany")
table(dana$EpisType[!is.na(dana$IDMDoutall)], useNA = "ifany")

```

We could use *IDMDoutrep* instead of *IDMDoutall* to match this only for the district the MEP represented.

5. Parties without ParlGov ID

Table 5. Parties without ParlGov ID

IDpty	pName
-9901	Sinyo Edinstvo Blue Unity
-9902	Régions et Peuples Solidaires
-9903	Europe citoyenne
-9904	Paese Nuovo
-9905	Strana demokratického Slovenska
-9906	Koaliziya KOD
-9907	Prasinoi
-9908	Scottish Socialist Party
-9909	Europees Verkiezers Platform Nederland
-9910	Forza Nuova
-9911	Scelta Europea
-9912	We Demand a Referendum Now
-9913	Io cambio - Movimento Associativo Italiani all’Estero
-9914	Libertas.cz
-9915	Artikel50
-9916	Traditional Unionist Voice
-9917	Parti de la France

6. Main sources of electoral data

Table 6. Main data sources of election results (links live as of summer 2021)

Country	Source
Multi-country	http://www.parlgov.org/ https://www.europarl.europa.eu/election/listcand/cceu01.htm https://www.europarl.europa.eu/elections2004/ep-election/sites/en/sitemap/
Austria	https://www.bmi.gv.at/412/Europawahlen/Europawahl_1999/start.aspx (replace year in 1999 link for 2004, 2009 and 2014)
Belgium	https://elections2004.belgium.be/fr/index.html (replace year in 2004 link for 2009 and 2014)
Bulgaria	https://results.cik.bg/
Croatia	https://www.izbori.hr/arhiva-izbora/index.html#/app/home
Cyprus	Personal communication with election authorities
Czech Republic	https://volby.cz/index_en.htm
Denmark	http://www.dst.dk/pukora/epub/Nyt/1999/NR253.pdf https://im.dk/media/7885/europa-parlamentsvalget-den-13-juni-2004.pdf http://www.oim.dk/media/15139/valget-til-europa-parlamentet-7-juni-2009.pdf http://valg.oim.dk/media/16116/saernummer-af-statistiske-efterretninger-om-valg-til-europa-parlamentet-25-maj-2014
Estonia	https://www.valimised.ee/en/archive/european-parliament-elections/european-parliament-elections
Finland	https://www.stat.fi/tk/he/vaalit/vaalit99euro/vaalitilastot_eu99_en.html https://www.stat.fi/til/euvaa/tau_en.html
France	https://www.interieur.gouv.fr/Elections/Les-resultats/Europeennes/elecresult__europeennes_2004/(path)/europeennes_2004 https://www.interieur.gouv.fr/Elections/Les-resultats/Europeennes/elecresult__europeennes_2009/(path)/europeennes_2009 http://www.interieur.gouv.fr/Elections/Les-resultats/Europeennes/elecresult__ER2014/(path)/ER2014/index.html
Germany	https://www.bundeswahlleiter.de/en/dam/jcr/0872e1f8-935a-45d6-a0f1-a3352fb4bc69/ew_ergebnisse_gesamt.pdf
Great Britain	http://researchbriefings.files.parliament.uk/documents/RP99-64/RP99-64.pdf http://researchbriefings.files.parliament.uk/documents/RP04-50/RP04-50.pdf https://en.wikipedia.org/wiki/1999_European_Parliament_election_in_the_United_Kingdom (change the four-digit year accordingly for later elections) https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-r
Greece	https://ekloges-prev.singularlogic.eu/may2014/e/public/index.html?lang=en#\{ %22cls%22:%22main%22,%22params%
Hungary	(closed list, nationwide district)
Ireland	http://electionsireland.org/results/europe/

Table 7. Main data sources of election results (continued)

Country	Source
Italy	http://elezionistorico.interno.it/index.php?tpel=E
Latvia	https://www.cvk.lv/cgi-bin/wdbcgiw/base/eiro.veles_rez04a.sak https://www.cvk.lv/cgi-bin/wdbcgiw/base/eiro9.veles9_rezult.sak http://ep2014.cvk.lv/?lang=1
Lithuania	http://www.vrk.lt/statiniai/puslapiai/rinkimai/2004/euro/rezultatai/rez.1.18.htm https://www.vrk.lt/statiniai/puslapiai/2009_ep_rinkimai/output_lt/rezultatai_daugiamand_apygardose/rezultatai_daugiamand_2009.htm (replace year in 2009 link for 2014)
Luxembourg	https://elections.public.lu/fr/elections-europeennes/1999/resultats/circonscriptions/unique/index.html#liste2 https://elections.public.lu/fr/elections-europeennes/2004/resultats-officiels/resultats-officiels.eu.2004.pdf https://elections.public.lu/content/dam/elections/fr/elections-europeennes/2009/resultats-officiels/recensement_general_2009.pdf https://elections.public.lu/fr/elections-europeennes/2014/resultats-officiels/RECENSEMENT_GENERAL_2014.pdf
Malta	http://www.um.edu.mt/projects/maltaelections/elections/mep
Netherlands	Preference votes from Protocols (Proces verbaal) of the Electoral Council (https://english.kiesraad.nl/)
Northern Ireland	https://www.eoni.org.uk/getmedia/7632ea3d-d26b-4d5e-b09d-d8d19f2480ed/european_election_results_1999 https://www.eoni.org.uk/Elections/Election-results-and-statistics/Election-results-and-statistics-2003-onwards/Election-results-and-statistics-2003-onwards-2003-2014 (replace year in 2004 link for 2009 and 2014)
Poland	http://pe2004.pkw.gov.pl/ http://pkw.gov.pl/591_Wyniki_glosowania/1/8767_Wyniki_glosowania_wg_okregow_wyborczych https://pe2009.pkw.gov.pl/PUE/EN/WYN/M/index.htm http://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20140000692/O/D20140692.pdf
Portugal	https://www.eleicoes.mai.gov.pt/europeias2004/concorrentes.html https://www.eleicoes.mai.gov.pt/europeias2009/candidatos.html (replace year in 2009 link for 2014)
Romania	(closed list, nationwide district)
Slovakia	https://volby.statistics.sk/ep/ep2004/ (replace year in 2004 link for 2009 and 2014)
Slovenia	https://www.dvk-rs.si/arhivi/ep2004/rez_kan.html (replace year in 2004 link for 2009 and 2014)
Spain	https://es.wikipedia.org/wiki/Anexo:Candidaturas_para_las_elecciones_al_Parlamento_Europeo_de_1999_(Espa%C3%B1a) https://es.wikipedia.org/wiki/Anexo:Candidaturas_para_las_elecciones_al_Parlamento_Europeo_de_2004_(Espa%C3%B1a)
Sweden	Statistiska centralbyrån (2001). Valet till Europaparlamentet 1999. Stockholm: SCB. Available at: http://share.scb.se/ov9993/data/historisk%20statistik//SOS%201911-/Valstatistiken/Valet%20till%20Europaparlamentet1999 http://web.archive.org/web/20040803085228/http://www.val.se/val/ep2004/resultat/protokoll/index.html http://web.archive.org/web/20100216195556/http://www.val.se/val/ep2009/slutresultat/protokoll_00E.pdf https://data.val.se/val/ep2014/slutresultat/protokoll/protokoll_00E.pdf