

The Comparative Study of Electoral Systems (CSES) Integrated Module Dataset (IMD)¹

Stimulus Paper

Stephen Quinlan

GESIS Leibniz Institute for the Social Sciences, Mannheim

Christian Schimpf

GESIS Leibniz Institute for the Social Sciences, Mannheim

David Howell

Centre for Political Studies, University of Michigan

Katharina Blinzler

GESIS Leibniz Institute for the Social Sciences, Köln

Slaven Zivkovic

GESIS Leibniz Institute for the Social Sciences, Köln

Bojan Todosijevic

Institute for the Social Sciences, Belgrade

December 4, 2018

¹ More information on the CSES Project and the Integrated Module Dataset (IMD) can be found at: www.cses.org. All authors are members of the CSES Secretariat as of December 4, 2018. The CSES Secretariat comprises the central staffing and operations for the CSES project, under the leadership of the Chair of the CSES Planning Committees. Support for the activities of the CSES Secretariat was received from several sources during the development of the CSES IMD. These include the American National Science Foundation (NSF) (grants SES-1420973 and grant SES-1760058). The CSES Secretariat activities at the GESIS Leibniz Institute for the Social Sciences, Germany were funded by GESIS. The Center for Political Studies at the University of Michigan provided additional support. We would like to express our gratitude to members of the CSES Module 5 Planning Committee, especially John Aldrich, its Chair, for their advice and suggestions on several issues in the development of CSES IMD. Earlier versions of this paper were presented to the CSES Planning Committee Meetings in Mannheim, Germany in November 2017 and in Göteborg, Sweden in June 2018, as well as the CSES Plenary Meeting at Harvard University in September 2018.

Summary

This Stimulus Paper provides an overview of the background to and key principles underlying the Comparative Study of Electoral Systems (CSES) Integrated Module Dataset (IMD). The document provides a brief overview of the evolution of the CSES project and why the CSES is now embarking on the development of an integrated data product. The remainder of the paper outlines the rationale underlying key decisions taken regarding CSES IMD, including the dataset's name and structure, the unifying principle of CSES IMD which specifies which variables and polities are eligible for inclusion, the principles underlying party/coalition numerical harmonization in CSES IMD, how CSES relational data (alphabetical coding) is handled in this new product and how the CSES IMD will be gradually developed across several different phases between 2018 and 2023.

1. An Integrated Harmonized CSES Dataset

1.1 The CSES project: a brief overview

The CSES was founded in 1994 and began collecting data in 1996. The project is a combined program of research among election study teams globally allowing for comparative analysis of electoral behavior worldwide. It involves participating polities including a standard module of survey questions in their post-election study. National election studies deposit these data with the CSES Secretariat, who harmonize the data to comparative standards and merge it with demographic, district and macro variables to create CSES data, which are then made public to users free of charge.

CSES data is cross-sectional (i.e., one point in time is studied in several countries) and with its integrated micro-macro design CSES data allows researchers to explore the role of context, both institutional and social factors to discover how they influence elections, values, and behavior. The studies run in module blocks of five years, with each module having a dedicated theme.² The project's governing board, the CSES Planning Committee, elected with input by the user community by participating national election studies for a five-year term, decides on a dedicated theme after submissions from the user community. As of December 2018, CSES has fielded four completed Modules across 55 polities with a fifth module in the field since 2016 and due for completion in 2021, and a sixth module in the early production phase.

² *Module 1: The Impact of Electoral Institutions on Citizens' Political Cognition and Behavior (1996-2002); Module 2: The Logic of Elections and Citizen Engagement and Cognition (2001-2006); Module 3: "Meaningful Choices" – Conditions where Elections provide a Meaningful Choice Set to Voters (2006-2011); Module 4: Distributional Politics, Social Protections, and Political Mobilization (2011-2016); Module 5: Democracy Divided? People, Politicians, and the Politics of Populism.*

1.2 The Road to a CSES Integrative Data Product

With four Modules now fielded, a fifth in the field, and a sixth module in the offing, the clamor for the CSES to produce an integrated data product has grown. In its 23rd year of data collection, over 30 polities have appeared in at least three of the four Standalone CSES Modules fielded, 63 percent of states have participated three times, and 13 election studies have run the Module five times or more. Hence, the possibilities for exploration of political behavior cross-nationally and over-time are enormous. Exploring phenomenon over time is an important and common task in political science, especially voting behavior, and has produced some trailblazing scholarly outputs (e.g., Campbell et al., 1960; Dalton, 2006; Heath et al., 1991; Lewis-Beck et al., 2008).

The creation of an over-time data product is well established in electoral research. National election studies in the Netherlands (Aarts & Todosijevic, 2009) and the United States (ANES, 2018) have responded by creating unified cross-sectional files. Cross-nationally, the European Voter Database (Thomassen et al., 2005) brought together election studies in six countries over-time (Britain, Denmark, Germany, the Netherlands, Norway, and Sweden) into a merged dataset, although these data were not based on common questions asked of each respondent, instead making a functional equivalence comparison (also see Thomassen, 2005). Regarding CSES, Giebler, Lichteblau, May, Melcher, Wagner, & Weßels (2016) from the WZB Berlin Social Science Centre broke new ground by creating the first publically available cumulative CSES dataset. This new departure brought together in one dataset CSES Modules 1, 2, and 3 and connected several variables of interest to scholars over time in a unified way for the first time.

There are several advantages to the CSES creating a unified and harmonized data product of its own. On the analytical front, it furthers science by allowing more users to explore behavior over time globally and potentially bringing context into focus in a new way. Such a product also facilitates trend analysis within one country over time. Additionally, pooled data such as these enables exploration of subgroups of citizens that are represented by few cases in a single, cross-section sample, but by many more cases when several samples are combined cross-nationally. All of this makes it more likely that CSES data will be used, leading to more scholarly output and return on investment in the project. It aids replication analyses as having a go-to CSES product ensures scholars are using the same consistent data source which will be archived and preserved but also which has been verified and checked thoroughly by the CSES Secretariat. Furthermore, it demonstrates that the CSES responds to user demand.³ Most

³ Some scholars have taken it upon themselves to merge together different aspects of Standalone CSES Modules for research with nearly 50 English language peer-reviewed research articles published based upon authors own cumulative efforts (e.g., Bingham Powell Jr, 2013; Blais, Guntermann, & Bodet, 2017; Ezrow, Homola, & Tavits, 2014; Helgason,

importantly, a new integrated product also opens up the possibilities of conquering new frontiers, specifically the challenge of harmonizing party coding and other demographic and macro data across CSES Modules.

In light of these considerations, the CSES Planning Committee (PC) mandated the CSES Secretariat to devise a Strategic Plan for the creation of a unified CSES data product incorporating harmonization of party/coalition code variables across Modules at its Meeting in Seattle in October 2015. In November 2017 upon delivery by the CSES Secretariat of the Strategic Plan to the Module 5 PC Meeting in Mannheim, the PC mandated the Secretariat to create an integrated harmonized data product encompassing CSES data from all modules to be known as the **CSES Integrative Module Dataset (IMD)**.

The remainder of this paper outlines the policies underpinning CSES IMD. We touch upon five key issues namely:

- Dataset name and structure
- The unifying principle of CSES IMD: The 3 and 1 Rule
- Harmonization of Party/Coalition Codes and Relational Data
- Introduction of pre-coded vote choice variables and Party/Coalition labels within data
- Product implementation

2. CSES IMD: Dataset name and structure

2.1 Dataset name

In identifying potential names for the integrated data product, we considered several issues. First, we wanted an easily understandable name which avoided confusion and ambiguity about the nature of the file, especially for non-native English speakers. Second, we wanted an original name and to avoid using a name already in use by similar products. Third, we wanted to ensure the name was market-friendly so to provide an identifiable brand of the project that would appeal to the user community and would be easily searchable on platforms such as Google. Fourth, we wanted a name which did not box in the dataset. We cannot predict the future direction of the CSES and whether its design will remain constant. Consequently, an adaptable name for the integrated data product was preferred.

Based on these criteria, we considered nine different names (or variations) for the name of the integrated data product. At the November 2017 CSES PC Meeting in Mannheim, the name

2016; Iversen & Soskice, 2015; McAllister, 2016; Sheppard, 2015). Please note: the aforementioned list of scholars is not meant to be comprehensive but rather illustrative.

CSES Integrated Module Dataset (IMD) was officially adopted as the integrated data product's name.

2.2 Data structure and variable names

The CSES convention regarding data structure has been that variables are ordered as follows: administrative; demographics; survey (micro level); macro level. As CSES users are familiar with this and it makes a clear delineation between different types of data, IMD retains this data structure.

Regarding the naming of variables, the CSES approach has been to assign each variable a five to ten alphanumeric identification (e.g., A1001 or D3006_LH_DC). The alphabetical character represents the CSES Module that the variable belongs to (for example variables beginning with 'A' refer to Module 1, variables beginning with 'B' refer to Module 2. The first digit refers to the group the variable belongs to.⁴ The remaining characters are numerical/alphabetical. The remaining digits/letters assigned to a variable name are arbitrary and refer to the order in which the variables appear in the dataset and/or the type of variable they represent etcetera).⁵

CSES IMD retains much of the existing CSES conventions. Instead of a single letter identifying the Standalone CSES Module, CSES IMD variables use the alphabetical character **IMD** (e.g., IMD1001). CSES IMD also follows the variable grouping structure in Standalone CSES Modules – i.e., variables beginning with a 1 refer to the administrative variables; variables beginning with a 2 refer to demographics etcetera.

3. The Unifying Principle of CSES IMD: *The 3 and 1 Rule*

The unifying principle of the CSES IMD is what variables and polities are eligible for inclusion in the dataset. Below, we outline the parameters we considered in deciding the unifying principle and then discuss the unifying principle itself.

3.1 Parameters conditioning the Unifying Principle

Our first consideration was the constraints the Standalone CSES Modules involve. As of 2018, CSES has been collecting data for 23 years inclusive with four modules completely fielded, a fifth currently in the field, and a sixth in the offing. If we focus on completed modules only, the

⁴ In Standalone CSES Modules, variables beginning with a 1 refer to the identification, weight, and study variables. Variables beginning with a 2 refer to demographic data. Variables beginning with a 3 refer to the micro-level survey data (the CSES module questionnaire). Variables beginning with a 4 refer to district-level data. Variables beginning with a 5 refer to macro-level data.

⁵ The ordering is somewhat arbitrary, although, for the micro-level component, it usually follows the order that questions are asked of respondents.

maximum number of data points at the variable level is four, and the modal number of observations at the polity level is 4.⁶ Further, as the CSES operates a themed module, many questions are module specific and have not been repeated across modules, thus reducing the number of variables repeated across time. Besides, at the polity level, while 35 polities have taken part in at least three CSES Modules, many (20 in total) have only fielded the questionnaire once or twice. Hence, there is a limit to the number of questions asked in repeated surveys and the number of consistent polity observations. That being said, with Module 5 in the field and Module 6 in production along with the classification of a core group of variables, we should observe more variables asked consistently over time, thus lessening the constraints.⁷

Our second consideration was what type of analyses CSES IMD would be used for. On the one hand, CSES IMD might be used for data maximization – i.e., adding more polities/elections and cases together to allow for more comprehensive analyses of a particular subject. With this approach, it is more about additional cases rather than a time component. Conversely, CSES IMD might be used to explore trends over time on several metrics cross-nationally or within polities. This brings into sharp focus what data would be necessary for a meaningful trend to be established. A trend refers to a general course, prevailing tendency, or a direction. It is represented by a gradual change in a condition with a series of data points moving in a particular direction over time. As a rule, the more data points we have, the more confident we can be about the observed trend. Yet, there is little agreement on how many data points we need to infer a trend, besides a consensus that more than two are expected. The challenge was thus to devise with a unifying principle that can accommodate both of these research approaches.

3.2 *The 3 and 1 Rule*

For eligibility for inclusion in the CSES IMD, all election study datasets had to be included in the published Standalone CSES Modules as they have been processed by the CSES Secretariat and thus gone through rigorous checking. The next step was to decipher a rule concerning the inclusion of variables and polities - *the 3 and 1 Rule*. The 3 applies to the variable level and stipulates that variables that have appeared three or more times in CSES modules 1-5 are eligible for inclusion in CSES IMD. This choice is informed by the idea that we cannot infer with any confidence a trend based on two data points. Thus, an inference regarding a trend requires more information. We maintain that the inclusion of variables that have appeared 3 times or

⁶ At the polity level, the range of observations is 1 to 7.

⁷ A core component of the CSES questionnaire was deciphered by the CSES Module 5 Planning Committee (see van der Brug, Cular, Just, Magalhães, & Oscarsson, 2016)

more thus strikes an apt balance between those who want to undertake trend analysis but also lends itself to those for who data maximization is the goal.

The 1 refers to the country level and proposes that all countries that have appeared in CSES are eligible for inclusion in the CSES IMD. This facilitates researchers who are interested in data maximization. Further, it may act as a greater incentive for a national election studies to include CSES in their questionnaire as not only will their data be included in the Standalone CSES Module but will also be part of the CSES IMD, heightening its exposure.

As CSES develops with more Standalone Modules, CSES IMD will evolve, and consequently, more and more variables will become eligible for inclusion and should soothe any concerns that the 3 and 1 Rule is too constricting. However, user instructions are provided in the CSES IMD Codebook detailing how practitioners can merge any data from Standalone CSES Modules and not included in CSES IMD themselves.

4. Harmonization of Party/Coalition and Leader Codes

Harmonization of party/coalitions across modules constitutes one of the most significant innovations about the CSES IMD. In this section, we first detail how the coding conventions applied in Standalone CSES Modules are unsuitable for harmonization and thus CSES IMD. Second, we detail how CSES IMD harmonizes parties/coalitions and deals with relational data. Third, we describe a set of rules applied to harmonization including how electoral alliances and party merges are dealt with.

4.1 CSES Party/Coalition/Leader coding conventions in Standalone CSES Modules

The CSES coding convention applied to Standalone Modules for parties, coalitions, and leaders grew organically from the beginning of the project and were a response to the relational data that CSES collects. This resulted in CSES adopting a dual coding scheme for these data: a numerical structure and an alphabetical structure. All the parties/coalitions and leaders participating in the current or the previous election receive a numerical code. Meanwhile, for relational data, the six most popular parties/coalitions, and their leaders are assigned alphabetical codes (A-F).⁸ The alphabetical codes are linked to the specific data collected on parties both at the micro and macro level including respondent placement of parties, respondent

⁸ To be eligible for an alphabetical code a party must have achieved at least 1% of the national vote share. Parties A-F are conventionally listed in descending order of their share of the popular vote in the parliamentary elections CSES uses three additional letters, G, H, and I, to denote supplemental parties. They may, but do not have to, according to with how parties A-F are ordered and are voluntarily provided by each country's election study and often reflect significant parties within a country (e.g., Plaid Cymru in Britain). CSES has tried to ensure that numerical codes and alphabetical codes have some interlink, with a new policy implemented for most of the studies in Module 4 that parties assigned numerical code 1 are assigned the Party A code, parties/coalitions assigned numerical code 2 are also assigned Party B code etc...

likeability of parties, and expert classification of party families. The full list of relational variables are detailed in Table 1.

Table 1 Variables in CSES Standalone Modules and for which the alphabetical coding is applied to

Type	Variable
Micro	Like-Dislike - Party A-I
	Like-Dislike - Leader A-I
	Left-Right - Party A-I
Macro	Percent Vote - Lower House - Party A-I
	Percent Seats - Lower House - Party A-I
	Percent Vote - Upper House - Party A-I
	Percent Seats - Upper House - Party A-I
	Percent Vote - President R1 - Party A-I
	Percent Vote - President R2 - Party A-I
	Ideological Family Parties A-I
Left-Right Parties A-I	

The above coding scheme means the numerical codes applied to parties/coalitions across Standalone CSES Modules have not been consistent, primarily because Standalone CSES Modules were following other conventions (see footnote 8) and party/coalition/leader coding was not thought of in a longitudinal way. Table 2 illustrates three examples. We see that in Germany, the Christian Democratic Union (CDU) has been assigned the numerical code 1 in Modules 1 and 3 but the numerical code 2 in Modules 2 and 4. In New Zealand, the New Zealand First Party was assigned the numerical code 3 for Modules 1 and 2 but numerical code 4 for Modules 3 and 4. Meanwhile, the Israeli Labor Party was assigned numerical code 2 for Modules 1-3 but numerical code 3 for Module 4. In sum, numerical party coding is inconsistent making it challenging and costly for users to analyze multiple CSES modules in tandem.

Table 2 Numerical codes assigned to three parties across Standalone CSES Modules 1-4

Party	Polity	M1	M2	M3	M4
Christian Democratic Union (CDU)	Germany	1	2	1 ⁹	2
HaAvoda (Labor Party – MHH)	Israel	2	2	2	3
New Zealand First (NZF)	New Zealand	3	3	4	4 ¹⁰

The second challenge relates to the alphabetical coding we have in CSES that covers the relational data. Figure 1 details the relational data structure about parties/coalitions and leaders – the main message: this data structure is complicated as respondents answers are linked to multiple parties, and thus the complications intensify in a unified dataset. Another challenge is that alphabetical codes are assigned inconsistently across the CSES Modules as the assignment of alphabetical codes is based on a module-specific rule. In addition, parties and coalitions appear inconsistently in the alphabetical codes, and thus there is a lack of uniformity across time. Hence, the challenge is to devise a means of allowing harmonization which takes into account the data structure CSES has. In exploring ways to harmonize party/coalitions across modules, the CSES Secretariat reviewed the harmonization of party codes in five cross-national projects which had a time-series dimension.¹¹ The Secretariat presented its findings in a memorandum to a Special General Meeting of the CSES Secretariat in October 2016 in Mannheim. What emerged from this analysis was that most projects had a unique and consistent numerical code for parties/coalitions. In most cases, this numerical code included some form of polity identifier within, although the choice of identifier was not consistent across projects.¹²

4.2 Numerical harmonization of parties/coalitions in CSES IMD

After extensive discussion at a Special General Meeting of the CSES Secretariat in Mannheim in October 2016, the CSES Secretariat, in conjunction with the CSES Planning Committee Chair John Aldrich, decided to harmonize parties/coalitions by numerical codes.

In CSES IMD, each party/coalition receives a unique numerical identifier that is consistent across modules. This seven-digit numerical identifier contains information on the country and a unique numerical value to distinguish the party/coalition. The first three digits of the identifier

⁹ 2005 study.

¹⁰ 2011 study.

¹¹ The projects we consulted were The Manifesto Project on Political Representation (Volkens et al., 2016), The European Election Study (EES), The Chapel Hill Expert Survey (Polk et al., 2017), The Eurobarometer Trend File 1970-2002 (Schmitt, Scholz, Leim, & Moschner, 2008), and The Constituency-Level Elections Archive (Kollman, Hicken, Caramani, Backer, & Lublin, 2016)

¹² For example CLEA uses the UN numerical country identifier while the CHES uses an arbitrary numerical system.

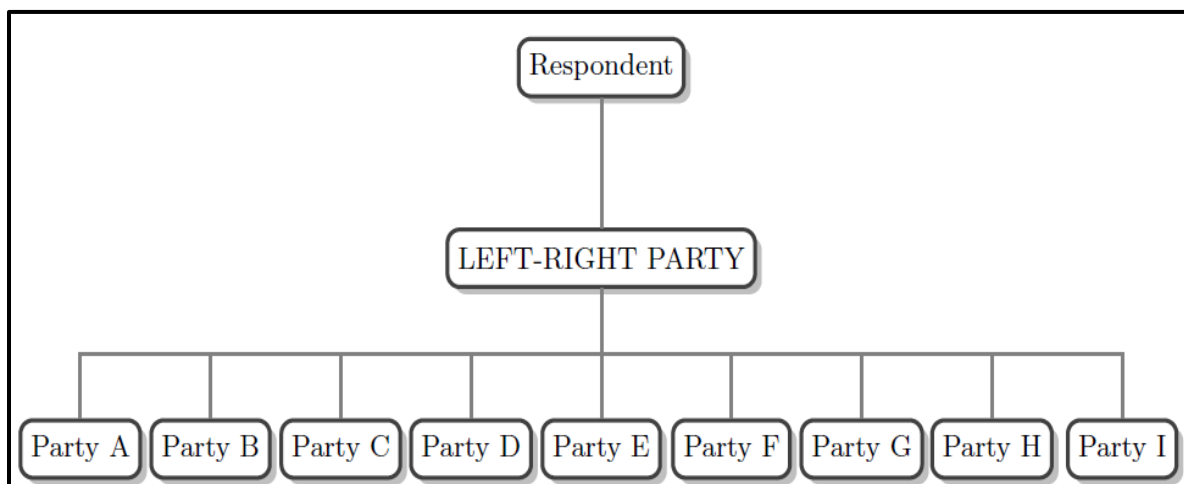


Figure 1 CSES relational data structure for CSES variables assigned alphabetical codes A-I. The example is based on the left-right placement of each party by a respondent.

consist of the three-digit UN Polity Identifier Code.¹³ The remaining four digits consists of numerical codes ranging from 0001 to 9999 with each party/coalition assigned a value that remains consistent across Modules.¹⁴ Table 3 details examples of this scheme applied to parties/coalitions from Germany and Australia. We highlight the country code part of the identifier in **GREEN** and the unique identifier assigned to each party in **ORANGE**.

In assigning of the last four digits, macro data specialists have assigned codes with consistent leading vote-getters in a polity being assigned lower values (for example in Table 3 users will notice the Liberal Party in Australia has been assigned the number 1 slot in Australia).

4.3 Additional conventions of numerical harmonization in CSES IMD

In case of a merger between two parties, the newly created party receives a different numerical code from its previous incarnations. We provide an example of this in Table 4 below. In 2009, the two Swiss parties, the Radical Democratic Party (FDP/PLR) and the Liberal Party

¹³ This variable is currently used in all CSES Modules – see variables A1006_UN, B1006_UN, C1006_UN, and D1006_UN and is also included in CSES IMD as IMD1006_UN.

¹⁴ We investigated the possibility that instead of reinventing the wheel with a set of new party/coalition numeric identifiers that we should instead adopt schemes used by other comparative projects. However, the cross-over in polities with other comparative projects and CSES is not exact. Moreover, each project has adopted schemas unique to its needs and all would pose complications for CSES. Hence, the development of a unique coding scheme for CSES IMD. However, as per CSES Module 4, the final version of CSES IMD will contain identifiers from popular comparative projects such as MARPOR/CMP and ParlGov to allow for easy data bridging between CSES IMD data and other comparative datasets.

Table 3 Proposed numerical coding of parties for Germany and Australia in the CSES IMD

Polity	CSES IMD Code	Party Name	Acronym
Germany	2760001	Christian Democratic Union/Christian Social Un.	Union
Germany	2760002	Christian Democratic Party	CDU
Germany	2760003	Christian Social Union in Bavaria	CSU
Germany	2760004	Social Democratic Party	SPD
Germany	2760005	Alliance 90/Greens	B90/GRUENE
Germany	Etc.	Etc.	Etc.
Australia	0360001	Liberal Party	LP
Australia	0360002	National Party of Australia	NPA
Australia	0360003	Australian Labor Party	ALP
Australia	0360004	Australian Democrats	AD
Australia	0360005	Australian Greens	AG
Australia	Etc.	Etc.	Etc.

(LPS/PLS), merged to form The Liberals (FDP/PLR). Both of the original parties keep their unique code (7560001 and 7560002 respectively). However, the newly formed party – the Liberals receives the code 7560011.

When parties compete in electoral alliances/coalitions, the alliance/coalition receives a different numerical code from the constituent parties that make it up. Table 5 provides an example with respect to Portugal. In 2015, the Social Democratic Party and the Popular Party competed together in an electoral alliance called “Portugal Ahead.” Each of the parties individually is assigned unique codes (6200001 and 6200002 respectively). However, the alliance/coalition formed for that election – the Portugal Ahead Alliance receives the code 6200014.

However, parties that merely undergo a name change will not receive a new unique code. For example, the Left Party in Germany (Die Linke) was known previously as the Party of Democratic Socialism (PDS) until it changed its name in 2005. However, the party is not assigned a new unique code. Instead, under Die Linke, you will find the results of the PDS in 1998 and 2002. Information about previous names is documented in the CSES IMD Codebook where possible.

Table 4 Proposed coding in case of a party merger in the CSES IMD:

Example for Switzerland

Polity	CSES IMD Code	Party Name	Acronym
Switzerland	7560001	Radical Democratic Party	FDP / PLR
Switzerland	7560002	Liberal Party	LPS / PLS
Switzerland	7560011	The Liberals	FDP / PLR

Table 5 Proposed coding in case of an electoral alliance in the CSES IMD:

Example for Portugal

Polity	CSES IMD Code	Party Name	Acronym
Portugal	6200001	Social Democratic Party	PSD
Portugal	6200002	Popular Party	CDS-PP
Portugal	6200014	Portugal Ahead (Electoral Alliance)	PSD - CDS-PP

4.4 How to deal with relational data (alphabetical codes) in the CSES IMD

Considering the decision to adopt harmonized numerical coding, the vexing question remained how to deal with the alphabetical coding and relational data. We explored three alternatives. The first would have been not to include any relational data in the CSES IMD. This was obviously impractical. Too much of the data of interest to practitioners is relational, so we ruled this out.

The second option was to abandon alphabetical classifications and attempt to harmonize alphabetical codes in a similar vein to numerical harmonization. For instance, instead of a variable LIKE-DISLIKE PARTY A, the dataset would include a variable LIKE-DISLIKE PARTY 1, ensuring harmonization with whatever party is coded 0001 for each polity. However, we rejected this idea because the number of extra variables needed to be included in the CSES IMD would need be equal to the maximum number of parties in a country that were ever assigned a letter multiplied by the number of variables that use the CSES alphabetical coding system. We estimate this would result in a dataset with a minimum of 110 extra columns, many of which would have empty cells. We feel this would result in a dataset too large and too complicated for users to handle. Crucially, this approach would be very resource intensive.

The third option and the one that CSES IMD adopts is to retain the alphabetical classifications assigned to parties/coalitions and leaders in Standalone CSES Modules, but to include identifiers with labels within CSES IMD indicating which party (and thus which leader) is assigned the A slot in a particular election in a polity. Figure 2 provides an example of how this looks in the dataset. We can see that in Australia, party 036003 The Australian Labor Party has been assigned the Party A code. A simple cross-tab with the election year in the polity allows users to quickly decipher which year this applies to (See Figure 3). For Australia, we can see the Labor Party was assigned as Party A for the election in 1996. This was replicated for each alphabetical assignment. Further, the CSES IMD codebook includes details allowing users to identify parties/coalitions and leaders corresponding to their A-I classification in each module.

We recognize that the above scheme for alphabetical coding does not constitute complete harmonization. However, bearing in mind the constraints of the data, taking into account resource, and not forgetting the genuine innovation of having numerical harmonization of parties/coalitions, this was considered to be the best option.

5. Other innovations in CSES IMD

Besides harmonization of party codes numerically across Modules, harmonization of CSES IMD includes other innovations. Among these is the inclusion of a pre-coded vote choice variable capturing whether a respondent voted for an incumbent or not. When it comes to vote choice at the individual level, the Standalone CSES Module provides users with vote choice variables separated by the type of election (Presidential, the lower house, the upper house) and the type of electoral system (vote for a party list, vote for a district candidate). This allows users significant flexibility to be able to transform the data to their own specifications.

CSES IMD retains this structure but introduces an innovation by including a pre-coded vote choice variable measuring incumbency. Drawing on the significant macro expertise of the Secretariat along with data provided by national collaborators, this dichotomous variable measures whether respondents voted for parties/candidates in government or for another party/candidate in the election under investigation. Studies using an incumbency variable are plentiful in political science, and this innovation is likely to save users significant effort in having to code this themselves. Thus, there is potential that this will stimulate more research. Moreover, having a variable such as this aids replication as there is now a go-to standard variable capturing incumbency.

PARTY A IDENTIFIER - NUMERICAL	Freq.	Percent	Cum.
0320001. ARG - Front for the Victory	1,406	0.50	0.50
0360001. AUS - Liberal Party	5,722	2.04	2.54
0360003. AUS - Australian Labor Party	3,671	1.31	3.84
0400001. AUT - Social Democratic Party	2,165	0.77	4.61
0560002. BEL - Socialist Party (Francop	1,960	0.70	5.31
0560004. BEL - Flemish Liberals and Dem	4,404	1.57	6.88
0760001. BRA - Workers Party	7,650	2.72	9.60
0760003. BRA - Party of the Brazilian D	1,000	0.36	9.95
1000001. BGR - Citizens for European De	999	0.36	10.31
1000019. BGR - National Movement for St	1,482	0.53	10.84
1120001. BLR - Communist Party of Belar	2,000	0.71	11.55
1240001. CAN - Liberal Party	7,727	2.75	14.30
1240002. CAN - Conservative Party	7,953	2.83	17.13
1520002. CHL - Independent Democratic U	2,400	0.85	17.98
1520005. CHL - Party for Democracy	1,173	0.42	18.40
1580001. TWN - Democratic Progressive P	3,845	1.37	19.77
1580002. TWN - Kuomintang of China	4,931	1.75	21.52
1910001. HRV - Croatian Democratic Unio	1,004	0.36	21.88
2030001. CZE - Czech Social Democratic	4,458	1.59	23.46
2030002. CZE - Civic Democratic Party	3,231	1.15	24.61
2080001. DNK - Social Democrats	2,001	0.71	25.32
2080002. DNK - Liberal Party	3,468	1.23	26.56
2330001. EST - Estonian Reform Party	1,000	0.36	26.91
2460001. FIN - Center Party	4,066	1.45	28.36
2460003. FIN - National Coalition Party	1,298	0.46	28.82
2500001. FRA - Union for a Popular Move	2,000	0.71	29.53
2500002. FRA - Socialist Party	2,014	0.72	30.25
2500008. FRA - Rally For The Republic	1,000	0.36	30.61
2760002. DEU - Christian Democratic Par	3,984	1.42	32.02
2760004. DEU - Social Democratic Party	7,060	2.51	34.53
3000001. GRC - Pan-Hellenic Socialist M	1,022	0.36	34.90
3000002. GRC - New Democracy	1,029	0.37	35.26
3000003. GRC - Coalition of the Radical	1,008	0.36	35.62
3440001. HKG - Democratic Alliance for	1,859	0.66	36.28
3440003. HKG - Democratic Party	2,256	0.80	37.09
3480001. HUN - Hungarian Socialist Part	2,725	0.97	38.06
3520001. ISL - Independence Party	6,151	2.19	40.25

Figure 2 Identification variable IMD5000_A allowing users to decipher which party has been allocated the Party A code in CSES over-time

Another step forward with CSES IMD is that with parties/coalitions now numerically harmonized, CSES IMD is able to include labels within the dataset for parties/coalitions (see Figure 2 above). This is likely to be a significant boon to users and reduce the potential for errors in coding.

PARTY A IDENTIFIER - NUMERICAL	ID COMPONENT - ELECTION YEAR							
	1996	1997	1998	1999	2000	2001	2002	2003
0320001. ARG - Front	0	0	0	0	0	0	0	0
0360001. AUS - Libera	0	0	0	0	0	0	0	0
0360003. AUS - Austra	1,798	0	0	0	0	0	0	0
0400001. AUT - Social	0	0	0	0	0	0	0	0
0560002. BEL - Social	0	0	0	1,960	0	0	0	0
0560004. BEL - Flemis	0	0	0	2,179	0	0	0	2,225
0760001. BRA - Worker	0	0	0	0	0	0	2,514	0
0760003. BRA - Party	0	0	0	0	0	0	0	0
1000001. BGR - Citize	0	0	0	0	0	0	0	0
1000019. BGR - Nation	0	0	0	0	0	1,482	0	0
1120001. BLR - Commun	0	0	0	0	0	1,000	0	0
1240001. CAN - Libera	0	1,851	0	0	0	0	0	0
1240002. CAN - Conser	0	0	0	0	0	0	0	0
1520002. CHL - Indepe	0	0	0	0	0	0	0	0
1520005. CHL - Party	0	0	0	1,173	0	0	0	0
1580001. TWN - Democr	0	0	0	0	0	2,022	0	0
1580002. TWN - Kuomin	1,200	0	0	0	0	0	0	0
1910001. HRV - Croati	0	0	0	0	0	0	0	0
2030001. CZE - Czech	0	0	0	0	0	0	948	0
2030002. CZE - Civic	1,229	0	0	0	0	0	0	0
2080001. DNK - Social	0	0	2,001	0	0	0	0	0
2080002. DNK - Libera	0	0	0	0	0	2,026	0	0
2330001. EST - Estoni	0	0	0	0	0	0	0	0
2460001. FIN - Center	0	0	0	0	0	0	0	1,196
2460003. FIN - Nation	0	0	0	0	0	0	0	0
2500001. FRA - Union	0	0	0	0	0	0	0	0
2500002. FRA - Social	0	0	0	0	0	0	0	0
2500008. FRA - Rally	0	0	0	0	0	0	1,000	0
2760002. DEU - Christ	0	0	0	0	0	0	0	0
2760004. DEU - Social	0	0	2,019	0	0	0	3,023	0
3000001. GRC - Pan-He	0	0	0	0	0	0	0	0
3000002. GRC - New De	0	0	0	0	0	0	0	0
3000003. GRC - Coalit	0	0	0	0	0	0	0	0
3440001. HKG - Democr	0	0	0	0	0	0	0	0
3440003. HKG - Democr	0	0	1,000	0	674	0	0	0
3480001. HUN - Hungar	0	0	1,525	0	0	0	1,200	0
3520001. ISL - Indepe	0	0	0	1,631	0	0	0	1,446

Figure 3 Identification variable IMD5000_A by Year allowing users to decipher which party/leader has been allocated the Party A code in a particular year

6. Product implementation: The evolution of CSES IMD

CSES IMD is an essential milestone for the CSES project and results in the CSES Secretariat taking responsibility for the production, management, and development of a new data product, which needs to be balanced with the Secretariat's other core responsibilities, namely the Standalone CSES Modules and project promotion and organization. Bearing this in mind, CSES IMD will be rolled out in a phased way with gradual releases taking place between 2018 and 2023. This implementation strategy is not only resource-friendly, but allows CSES IMD to be delivered to users quicker, is low risk from a data management perspective as concentration can be given to complex issues one by one, and allows feedback from users based on Advance Releases to be implemented in the Full Release. This strategy was endorsed by the CSES PC Meeting in Mannheim in November 2017.

References

- Aarts, C., & Todosijevic, B. (2009). Nationaal Kiesersonderzoek (NKO) 1971-2006 cumulative file. <https://doi.org/10.17026/dans-zcz-csbx>
- ANES. (2018). ANES Time Series Cumulative Data File (1948-2016). Retrieved from <https://electionstudies.org/project/anes-time-series-cumulative-data-file/>
- Bingham Powell Jr, G. (2013). Representation in Context: Election Laws and Ideological Congruence Between Citizens and Governments. *Perspectives on Politics*, 11(1), 9–21. <https://doi.org/10.1017/S1537592712003635>
- Blais, A., Guntermann, E., & Bodet, M. A. (2017). Linking Party Preferences and the Composition of Government: A New Standard for Evaluating the Performance of Electoral Democracy. *Political Science Research and Methods*, 5(2), 315–331. <https://doi.org/10.1017/psrm.2015.78>
- Campbell, D. E., Converse, Miller, & Stokes, D. (1960). *The American Voter*. New York, NY: Jon Wiley & Sons.
- Dalton, R. J. (2006). *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies* (Fourth Edition). Washington D.C.: CQ Press.
- Ezrow, L., Homola, J., & Tavits, M. (2014). When Extremism Pays: Policy Positions, Voter Certainty, and Party Support in Postcommunist Europe. *Journal of Politics*, 76(2), 535–547. <https://doi.org/10.1017/S0022381613001461>
- Giebler, H., Lichteblau, J., May, A., Melcher, R., Wagner, H., & Weßels, B. (2016). CSES Module 1-3 Harmonized Trend File [dataset]. May 31, 2016 version. <https://doi.org/10.7804/cses.trendfile.2016-05-31>
- Heath, A., Jowell, R., Curtice, J., Evans, G., Field, J., & Witherspoon, S. (1991). *Understanding political change: The British voter 1964-1987*. Pergamon Press.
- Helgason, A. F. (2016). Income-based voting and polarization over redistribution under alternative electoral systems. *Electoral Studies*, 42, 22–32. <https://doi.org/10.1016/j.electstud.2016.01.005>
- Iversen, T., & Soskice, D. (2015). Information, Inequality, and Mass Polarization: Ideology in Advanced Democracies. *Comparative Political Studies*, 48(13), 1781–1813. <https://doi.org/10.1177/0010414015592643>
- Kollman, K., Hicken, A., Caramani, D., Backer, D., & Lublin, D. (2016). Constituency-level elections archive [data file and codebook]. Ann Arbor, MI: Center for Political Studies, University of Michigan [producer and distributor]. *Ann Arbor: Centre for Political Studies, University of Michigan*. Retrieved from <http://www.electiondataarchive.org>
- Lewis-Beck, M. S., Jacoby, W. G., Norporth, H., & Weisberg, H. F. (2008). *The American voter revisited*. University of Michigan Press.
- McAllister, I. (2016). Democratic Consolidation in Taiwan in comparative perspective. *Asian Journal of Comparative Politics*, 1(1), 44–61. <https://doi.org/10.1177/2057891115620683>
- Polk, J., Rovny, J., Bakker, R., Hooghe, L., Jolly, S., Koedam, J., ... Zilovic, M. (2017). Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data. *Research & Politics*, 1–9.
- Schmitt, H., Scholz, E., Leim, I., & Moschner, M. (2008). The Mannheim Eurobarometer Trend File 1970-2002 (ed. 2.00). *GESIS Data Archive, Cologne (ZA3521)*. <https://doi.org/10.4232/1.10074>
- Sheppard, J. (2015). Compulsory voting and political knowledge: Testing a “compelled engagement” hypothesis. *Electoral Studies*, 40, 300–307.
- Thomassen, J. (Ed.). (2005). *The European Voter: a comparative study of modern democracies*. Oxford: Oxford University Press.
- Thomassen, J., Anderson, J., Goul Anderson, J., Klingemann, H. D., Schmitt, H., Wessels, B., ... Kumlin, S. (2005). The European Voter Project. Retrieved from

<https://www.gesis.org/en/services/data-analysis/more-data-to-analyze/more-international-data/the-european-voter-project/>

van der Brug, W., Cular, G., Just, A., Magalhães, P., & Oscarsson, H. (2016). *CSES Module 5 Core & Demographics Subcommittee Report*. Philadelphia, PA. Retrieved from http://www.cses.org/plancom/module5/CSES5_CoreDemographicsSubcommittee_FinalReport.pdf

Volkens, A., Lehmann, P., Theres, M., Merz, N., Regel, S., & Wessels, B. (2016). The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). *Berlin: Wissenschaftszentrum Berlin Für Sozialforschung (WZB)*.