



**digital
society
project**

Digital Society Survey

Codebook

Valeriya Mechkova (University of Gothenburg)
Daniel Pemstein (North Dakota State University)
Brigitte Seim (University of North Carolina, Chapel Hill)
Steven Wilson (Brandeis University)

March 2024



Table of Contents

1	Introduction	2
1.1	Funders	2
1.2	Cautionary Notes	2
1.3	Variable Types	3
1.4	Variable Versions and Suffixes	3
1.5	Aggregation	7
1.6	Variable Information	8
1.7	Suggested Citation	10
1.8	Country Units	12
1.9	Identifier Variables in the V-Dem and DSP Datasets	16
1.9.1	Country Name (country_name)	16
1.9.2	V-Dem Country ID (country_id)	16
1.9.3	Country Name Abbreviation (country_text_id)	16
1.9.4	Year (year)	16
1.9.5	Historical Date (historical_date)	16
1.9.6	Start of Coding Period (codingstart)	16
1.9.7	Gap in Coding Period Starts (gapstart)	16
1.9.8	Gap in Coding Period Ends (gapend)	16
1.9.9	Gap index (gap_index)	17
1.9.10	End of Coding Period (codingend)	17
1.9.11	COW Code (COWcode)	17
2	Digital Society Survey	18
2.1	Coordinated Information Operations	19
2.1.1	Government dissemination of false information domestic (C) (v2smgovdom)	19
2.1.2	Government dissemination of false information abroad (C) (v2smgovab)	19
2.1.3	Party dissemination of false information domestic (C) (v2smpardom)	20
2.1.4	Party dissemination of false information abroad (C) (v2smparab)	20
2.1.5	Foreign governments dissemination of false information (C) (v2smfordom)	21
2.1.6	Foreign governments ads (C) (v2smforads)	21
2.2	Digital Media Freedom	22
2.2.1	Government Internet filtering capacity (C) (v2smgovfilcap)	22
2.2.2	Government Internet filtering in practice (C) (v2smgovfilprc)	22
2.2.3	Government Internet shut down capacity (C) (v2smgovshutcap)	22
2.2.4	Government Internet shut down in practice (C) (v2smgovshut)	23
2.2.5	Government social media shut down in practice (C) (v2smgovsm)	23
2.2.6	Government social media alternatives (C) (v2smgovsmalt)	24
2.2.7	Government social media monitoring (C) (v2smgovsmmon)	24
2.2.8	Government social media censorship in practice (C) (v2smgovsmcenprc)	25
2.2.9	Government cyber security capacity (C) (v2smgovcapsec)	25
2.2.10	Political parties cyber security capacity (C) (v2smpolcap)	26
2.3	State Internet Regulation Capacity and Approach	26
2.3.1	Internet legal regulation content (C) (v2smregcon)	26
2.3.2	Privacy protection by law exists (C) (v2smprivex)	26
2.3.3	Privacy protection by law content (C) (v2smprivcon)	27
2.3.4	Government capacity to regulate online content (C) (v2smregcap)	27
2.3.5	Government online content regulation approach (C) (v2smregapp)	28



2.3.6	Defamation protection (C) (v2smlawpr)	28
2.3.7	Abuse of defamation and copyright law by elites (C) (v2smdefabu)	28
2.4	Online Media Polarization	29
2.4.1	Online media existence (C) (v2smonex)	29
2.4.2	Online media perspectives (C) (v2smonper)	29
2.4.3	Online media fractionalization (C) (v2smmefra)	30
2.5	Social Cleavages	30
2.5.1	Online harassment groups (C) (v2smhargr)	30
2.5.2	Other online harassment groups (C) (v2smhargrtxt)	31
2.5.3	Use of social media to organize offline violence (C) (v2smorgviol)	31
2.5.4	Average people's use of social media to organize offline action (C) (v2smorgavgact)	31
2.5.5	Elites' use of social media to organize offline action (C) (v2smorgelitact)	32
2.5.6	Types of organization through social media (C) (v2smorgtypes)	32
2.5.7	Other types of organization through social media (C) (v2smorgtypestxt)	33
2.5.8	Party/candidate use of social media in campaigns (C) (v2smcamp)	33
2.5.9	Arrests for political content (C) (v2smarrest)	33
2.5.10	Polarization of society (C) (v2smpolsoc)	34
2.5.11	Political parties hate speech (C) (v2smpolhate)	34

3 Appendix: Glossary



1 Introduction

The Digital Society Survey, designed by the Digital Society Project, contains questions pertaining to the political environment of the internet and social media. These data, which we collected using expert-coded surveys, provide information on topics related to coordinated information operations, digital media freedom, online media polarization, social cleavages, and state internet regulation capacity and approach.

For more information, please visit <http://www.digitalsocietyproject.org>

1.1 Funders

The Varieties of Democracy Institute (V-Dem) provided the use of its infrastructure for this data collection project. The V-Dem data team processed this survey using the standard V-Dem measurement modeling and quality control processes, using the V-Dem Institute infrastructure for collection and aggregating expert-based data on democracy, reaching out to a network of over 3,200 scholars from more than 180 countries. To learn more about V-Dem, and its funders, please visit: <https://www.v-dem.net/>.

In addition, DSP received support from Facebook to cover the costs of initial data collection. The National Science Foundation provided support (Grant No. SES-1423944) for the development of the methodological tools upon which we rely.

For data enquires: contact@digitalsocietyproject.org

1.2 Cautionary Notes

Both DSP and V-Dem are firmly committed to full transparency and data sharing. We ask users to take the following cautions into consideration when using the dataset.

- The *V-Dem Methodology* assumes five or more coders for the "contemporary" period. DSP spans a subset of this period, covering 2000-2023. We urge users to exercise caution when working with observations that were rated by fewer than five experts, and to pay careful attention to estimates of uncertainty around the point estimates that we provide for each observation. We strongly advise against using point estimates for country-variable-years with three or fewer (≤ 3) ratings. We suggest filtering these observations out before conducting any type of analysis. For this purpose, a special count-variable for each Country-Expert coded variable, which is suffixed with "_nr", is included in the dataset.
- These variables had issues with convergence: v2smforads and v2smgovsmalt. Please see individual codebook entries for additional information. For details on interpreting convergence information, please refer to 1.6, the Methodology Document and Pemstein et al. (2024).



1.3 Variable Types

The DSP data is gathered following the V-Dem rules and procedures. While V-Dem contains a number of types of data, all DSP variables are what V-Dem calls 'type C' variables:

- **Type C: Variables coded by Country Experts**

A Country Expert is typically a scholar or professional with deep knowledge of a country and of a particular political institution. Furthermore, the expert is usually a citizen or resident of the country. Multiple experts (usually 5 or more) code each variable. More information about the Country Experts can be found in the *V-Dem Methodology* document.

1.4 Variable Versions and Suffixes

Just like the V-Dem Dataset, the DSP data set contains several versions of the variables coded by country experts (type C variables).

- **Model Estimates**

"Model Estimates" — Measurement Model Output:

This version has no special suffix (*e.g.* v2elmulpar). This version of the variables provides country-year (country-date in the alternative dataset) point estimates from the V-Dem measurement model (see Pemstein et al. 2022). The measurement model aggregates the ratings provided by multiple country experts and, taking disagreement and measurement error into account, produces a probability distribution over country-year scores on a standardized interval scale (see the *V-Dem Methodology* document). The point estimates are the median values of these distributions for each country-year. The scale of a measurement model variable is similar to a normal ("Z") score (*e.g.* typically between -5 and 5, with 0 approximately representing the mean for all country-years in the sample) though it does not necessarily follow a normal distribution. For most purposes, these are the preferred versions of the variables for time series regression and other estimation strategies.

"Model Estimates Measure of Uncertainty" — Measurement Model Highest Posterior Density (HPD) Intervals:

This version has the suffixes: "codelow" and "codehigh" (*e.g.* v2elmulpar_codelow and v2elmulpar_codehigh). These two kinds of variables ["code low" and "code high"] demarcate the interval in which the measurement model places 68 percent of the probability mass for each country-year score, which is approximately equivalent to one standard deviation upper and lower bounds. If the underlying posterior distribution is skewed, the HPDs reflect this with unequal distances between the point estimate and the high and low estimates. We also provide a standard calculation for standard deviation which is marked with the suffix "sd" (*e.g.*, v2elmulpar_sd). The SD might be used to compute the standard frequentist confidence intervals.

- **Original Scale (*_osp)**

"Original Scale" — Linearized Original Scale Posterior Prediction:

This version has the suffix "_osp," (*e.g.* v2elmulpar_osp). In this version of the variables, we have linearly translated the measurement model point estimates back to the original ordinal scale of each variable (*e.g.* 0-4 for v2elmulpar_osp) as an interval measure. The decimals in the _osp version roughly indicate the distance between the point estimate from the linearized measurement model posterior prediction and the threshold for reaching the next level on the original ordinal scale. Thus, a _osp value of 1.25 indicates that the median measurement model posterior predicted value was closer to the ordinal value of 1 than 2 on the original scale. Technically, it calculates the sum of the posterior probabilities that the estimate is in a particular category: If a particular country-year-variable has a probability of 90% to be in category "4", a 10% probability of being in category "3", and 0% probability of being in categories "2", "1", and "0", the result is a value of 3.9 ($4 * 0.9 + 3 * 0.1 = 3.6 + 0.3$). Since there is no conventional theoretical justification for linearly mapping ordinal posterior predictions onto an interval scale, these scores should primarily be used for heuristic purposes. Using the "Ordinal Scale" estimates—or incorporating the properties of ordinal probit models into the estimation procedure—is thus preferable to using the _osp estimates in statistical



analyses. However, since the `_osp` version maps onto the coding criteria found in the V-Dem Codebook, and is strongly correlated with the Measurement Model output (typically at .98 or higher), some users may find the `_osp` version useful in estimating quantities such as marginal effects with a clear substantive interpretation. If a user uses `_osp` data in statistical analyses it is imperative that she confirm that the results are compatible with estimations using Measurement Model output.

"Original Scale Measure of Uncertainty" — Linearized Original Scale HPD Intervals:

This version has the suffixes "codelow" and "codehigh" (e.g. `v2elmulpar_osp_codelow` and `v2elmulpar_osp_codehigh`). We estimate these quantities in a similar manner as the Measurement Model Highest Posterior Density Intervals. These two variables ["code low" and "code high"] demarcate the interval in which the measurement model places 70 percent of the probability mass for each country-year score, which is approximately equivalent to one standard deviation upper and lower bounds. If the underlying posterior distribution is skewed, the HPDs reflect this with unequal distances between the point estimate and the high and low estimates. We also provide a standard calculation for standard deviation which is marked with the suffix "sd" (e.g., `v2elmulpar_sd`). The SD might be used to compute the standard frequentist confidence intervals.

- **Ordinal Scale (*`_ord`)**

"Ordinal Scale" — Measurement Model Estimates of Original Scale Value:

This version has the suffix "`_ord`" (e.g. `v2elmulpar_ord`). This method translates the measurement model estimates back to the original ordinal scale of a variable (as represented in the Codebook) after taking coder disagreement and measurement error into account. More precisely, it represents the most likely ordinal value on the original codebook scale into which a country-year would fall, given the average coder's usage of that scale. More specifically, we assign each country-year a value that corresponds to its integerized median ordinal highest posterior probability category over Measurement Model output.

"Ordinal Scale Measure of Uncertainty" — Original Scale Value HPD Intervals:

This version has the suffixes - "codelow" and "codehigh" (e.g. `v2elmulpar_ord_codelow` and `v2elmulpar_ord_codehigh`). We estimate these values in a similar manner as the Measurement Model Highest Posterior Density Intervals. These two variables ["code low" and "code high"] demarcate the interval in which the measurement model places 70 percent of the probability mass for each country-year score, which is approximately equivalent to one standard deviation upper and lower bounds. If the underlying posterior distribution is skewed, the HPDs reflect this with unequal distances between the point estimate and the high and low estimates. We also provide a standard calculation for standard deviation which is marked with the suffix "sd" (e.g. `v2elmulpar_sd`). The SD might be used to compute the standard frequentist confidence intervals.

- **Comparison of Model Estimates, Original Scale, and Ordinal Scale**

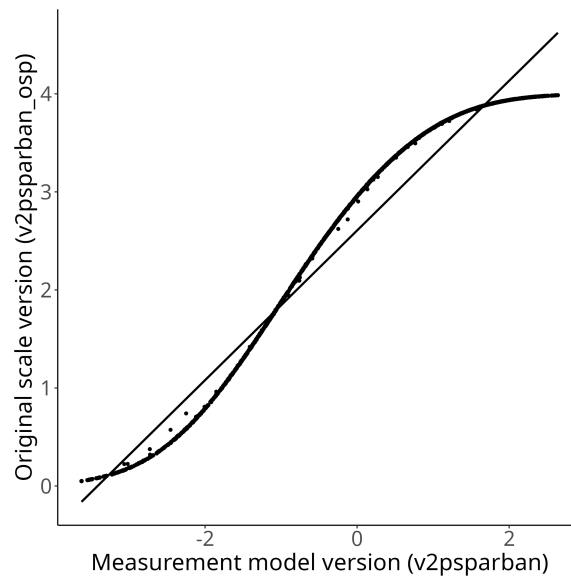
Each of these versions of the data emphasizes a different aspect of the same underlying estimates. The measurement model direct output "Model Estimates" prioritize measuring concepts on an equal-interval scale so that, for example, an increase from -2.7 to -1.7 is a change in magnitude that is equivalent to an increase from 0.7 to 1.7: both are an increase of one point. The same goes for the Bayesian confidence intervals that accompany each variable. Interval-level measurement makes this version the only one that is, strictly speaking, appropriate for analyses such as ordinary regression that are designed for continuous variables. The downside is that the scale of the measurement model is different from the scale used in the codebook and translating from the "Model Estimates" to codebook values must be done on an indicator-by-indicator basis. It is therefore difficult to substantively interpret "Model Estimates," or marginal effects based on these estimates, in terms of the codebook scale.

The "Original Scale" (`_osp`) versions rescale the "Model Estimates" to cover the range of the discrete original scale scores in the codebook (commonly 0 to 3, 4, or 5) in order to make it easier for users to see where the continuous scores are located on the category definitions in the codebook; be able to make descriptive analyses that are interpretable in substantive terms; and supplement regression-type analyses with substantive interpretations and estimates. However, a consequence is that these estimates distort the intervals to some extent (varying between variables). Despite the very high correlation between the measurement



model version and the original scale version, a unit change at one range of the scale is not strictly equivalent to a unit change at a different range. For most variables, the original scale version is a non-linear S-curve transformation of the measurement model estimates because the *_osp transformation tends to expand the intervals close to the median and shrink the intervals at the extremes. The confidence intervals are similarly distorted. There also tend to be a very small number of cases that deviate minutely from this pattern. The shape of the S-curve varies somewhat across variables. Figure 1 illustrates this relationship using the “Party ban” variable.

Figure 1: Comparison of original scale and equal-interval measurement model estimates for Party ban (all countries, 1900-2023)

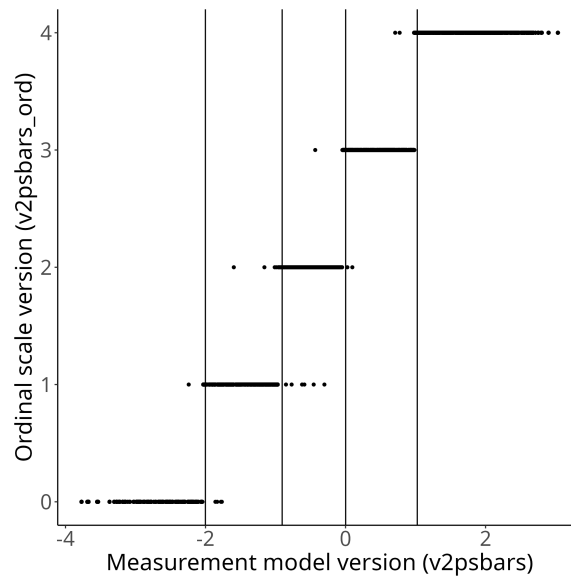


Note: Diagonal line shows what the linear relationship would be.

The “Ordinal Scale” version of the variables favors a one-to-one correspondence between the codebook definitions and ordinal-level scores, by converting continuous score into the most likely (highest posterior density) ordinal codebook value. This conversion offers a categorical interpretation for each data point. However, it does so by omitting the degrees of variation found in the model estimates and the original scale version. Figure 2 clarifies the relationship between the ordinal version and model estimates using the example of Barriers to political parties (v2psbars). Clearly each ordinal category corresponds to a range of continuous values, and some continuous values fall entirely outside of the categories. Observations that are at the low end of category 2 may be more similar to those at the high end of category 1 than they are to those at the high end of category 2. The thresholds come from estimates in the measurement model but are somewhat poorly identified (i.e. there is a relatively high degree of uncertainty around them), and there is therefore not always a perfect correspondence between the continuous ranges and ordinal categories. For instance, some observations that are assigned to category 1 have continuous scores in the same range as most of the scores in category 2, and so on. The confidence bounds for the ordinal scores are similarly discrete, typically one point above or below the score, although sometimes they are the same as the score.



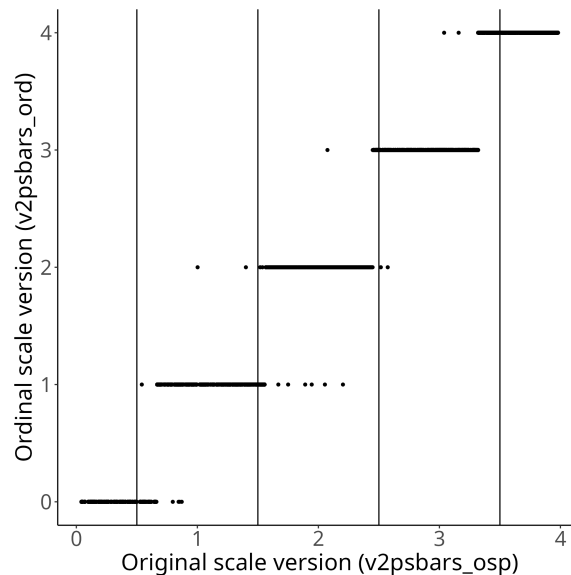
Figure 2: Comparison of ordinal scale and equal-interval measurement model estimates for Barriers to political parties (all countries, 1900-2023)



Note: Vertical bars added to suggest thresholds between categories.

The relationships are similar between the original scale and ordinal scale versions, as shown in Figure 3, although they suggest that the most natural thresholds between ordinal scores are not necessarily midway between them.

Figure 3: Comparison of original scale and equal-interval measurement model estimates for Barriers to political parties (all countries, 1900-2023)

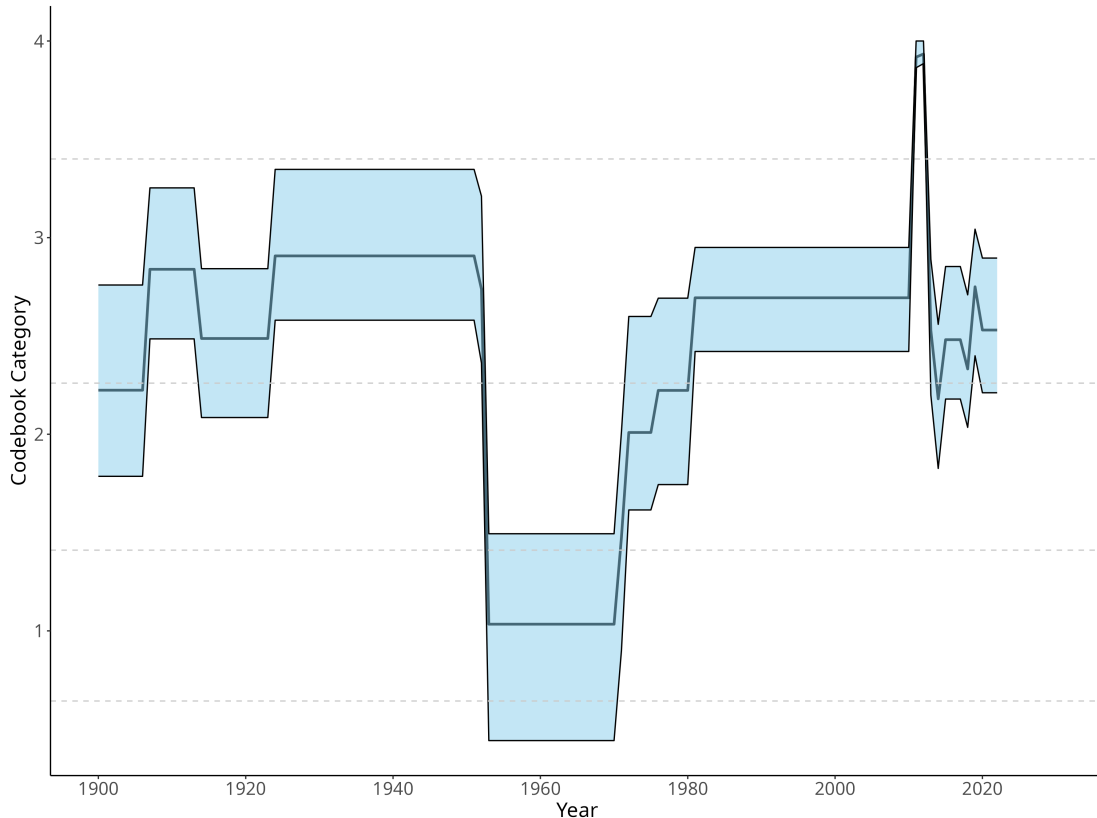


Note: Vertical bars added to suggest thresholds between categories.

No single version of our C variables captures all the relevant information – an equal-interval continuous score and the corresponding ordinal score that is easily interpreted. However, all of this information is displayed in the variable-specific line graphs on the V-Dem website. As shown in the example in Figure 4, this visualization plots the equal-interval measurement model estimates (with their corresponding 68 percent Bayesian confidence intervals) for Egypt by year,

on horizontal gridlines that divide the ranges that correspond to the ordinal scores, which are the y-axis labels. Note that the intervals between thresholds vary in size; this reflects the fact that the distance between a 3 and a 4 is greater than the distance between a 1 and 2, while unit changes are comparable along all values on the scale. The gridlines are approximations that sometimes do not have a one-to-one relationship with the measurement model estimates, but they are the same thresholds (specifically, the average global thresholds used in the measurement model) that are used to calculate the ordinal and original scale versions of the data.

Figure 4: Measurement model estimates for Party Ban (v2psparban) in Egypt (1900-2022)



- **Number of Coders per Country, Variable and Year/Date (*_nr)**

The number of V-Dem Country Experts (regular coders, bridge- and lateral coders) who provided data on country, variable and year. V-Dem's methodology is based on the assumption that we have a minimum of five Country Experts for every single country-variable-year. Sometimes, however, we end up with fewer than five Country Experts. From v7 of the Country-Year, and the Country-Date type datasets, we provide all data we have for full transparency. By providing the number of Country Experts for each country-variable-year/date, we suggest that users primarily base analyses on observations based on five or more coders. We strongly advise against using observations based on three or fewer coders. This concerns all C type variables.

- **Arithmetic Mean of Coder Answers per Country-Year (*_mean)**

It is commonplace to aggregate respondents' data to the level of country or country-year using arithmetic mean in order to merge it with other country-level data. V-Dem Institute provides such variables for every expert-coded variable aggregated by the Measurement Model in Country-Date/Year dataset.

1.5 Aggregation

C-variables, ratio/percentage variables, and High-Level/Mid-Level Democracy indices are aggregated from the country-date level to the country-year level by the day-weighted mean.



1.6 Variable Information

The following information is available per variable (if applicable):

Additional versions : Indicates if the variable is also available in the following versions; *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean and/or *_nr. Detailed information about the different versions can be found in section 1.4 (*Variable Versions and Suffixes*).

Available versions: Lists the available variable types (Only applicable for ordinalized versions of indices).

Question: The question that the variable attempts to measure.

Clarification: Definition of key terms, clarification of scope-conditions, contexts, and any other features needed to understand the question (if any). All key terms appear in the Glossary (*Appendix*), unless they are specific to a single section (in which case they only appear in the introduction to that section or in the clarifications for particular questions). Key terms are sometimes cross-referenced with hypertext.

Responses: Numeric, Percentage, Text, Date, Countries, or specific response categories (listed below under "Answer-types" and "Scales").

Answer-Types:

Multiple-choice: Where a coder can select only one answer. This is the usual protocol and is therefore not noted.

Multiple-selection: Where a coder can select more than one answer. For most multiple-selection variables, the dataset contains both the original variable as well as a set of dummies for each of the responses.

Ordering (only applicable to a selection of C variables): This relates to the ordering of questions when the coding of one indicator depends upon the coding of other indicators (*i.e.*, whenever there is some alteration of the serial ordering of questions as listed in this document).

Aggregation (only applicable to indices): Explanation of how an index is constructed.

Scale: Dichotomous, Nominal, Ordinal, or Interval/Ratio (Extra response options such as N/A or Other, are not counted as part of this classification).

Cross-Coder Aggregation (only applicable to C variables): IRT, Bayesian ordinal item response theory measurement model (see the *V-Dem Methodology* document). Available in mode and mean.

Cleaning: Specifies if observations are set to missing based on values on other variables.

Citation: Suggested citation when using the specific variable.

Convergence: V-Dem assesses convergence among expert-coded variables using the Gelman-Rubin Diagnostic. Specifically, we consider a variable to have converged if no more than 5% of parameters in each of the relevant parameter sets (universal thresholds, main-country-coded thresholds, expert thresholds, expert reliability, and country-date latent trait estimates) has $\hat{R} \geq 1.01$. We assess BFA convergence in a similar manner across relevant model parameter sets (intercept, slope, measurement standard error, and country-date latent trait estimates), but using $\hat{R} \geq 1.1$. We provide convergence information for a given variable only if a set of model parameters did not converge, reporting these set(s). Note that if country-date latent trait estimates converged (*i.e.* we do not mention them in the convergence details) it means that the convergence issues likely reflect a problem with model parameter identification, and the latent trait estimates are relatively safe.

Years: Available coverage for the respective variable. For more information on country-specific year



coverage, see *the country table*.

Note: Additional information about the variable.



1.7 Suggested Citation

Nota bene: If a variable drawn from the DSP dataset plays an important role in your project (published or unpublished), please cite the following:

- **DSP Dataset:**
Mechkova, Valeriya, Daniel Pemstein, Brigitte Seim, and Steven Wilson. 2024. "DSP Dataset v6" Digital Society Project (DSP).
- **DSP Introduction:**
Mechkova, Valeriya, Daniel Pemstein, Brigitte Seim, and Steven Wilson. 2019. "Measuring Internet Politics: Introducing the Digital Society Project" Digital Society Project (DSP).
- **V-Dem Methodology:**
Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Kyle L. Marquardt, Juraj Medzihorsky, Natalia Natsika, Daniel Pemstein, Linnea Fox, Lisa Gastaldi, Ana Good God, Sandra Grahn, Josefine Pernes, Oskar Rydén, Johannes von Römer, Eitan Tzelgov, Yi-ting Wang, and Steven Wilson. 2024. "V-Dem Methodology v14" Varieties of Democracy (V-Dem) Project.
- **V-Dem Measurement Model:**
Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri, and Johannes von Römer. 2024. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data". V-Dem Working Paper No. 21. 9th edition. University of Gothenburg: Varieties of Democracy Institute.

Given the DSP's intellectual debt to the larger V-Dem project, we strongly encourage users of the DSP data to also include some or all of the following citations:

- **V-Dem Dataset:**
Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Natalia Natsika, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Oskar Rydén, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Yi-ting Wang, Tore Wig, Steven Wilson and Daniel Ziblatt. 2024. "V-Dem [Country-Year/Country-Date] Dataset v14" Varieties of Democracy (V-Dem) Project.
<https://doi.org/10.23696/mcwt-fr58>.
- **V-Dem Codebook:**
Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Oskar Rydén, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, and Daniel Ziblatt. 2024. "V-Dem Codebook v14" Varieties of Democracy (V-Dem) Project.
- **V-Dem Country Coding Units:**
Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Lisa Gastaldi, Ana Good God, and Sandra Grahn. 2024. "V-Dem Country Coding Units v14" Varieties of Democracy (V-Dem) Project.



- **V-Dem Organization and Management:**

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Susanna Burmeister, Linnea Fox, Lisa Gastaldi, Sandra Grahn, Natalia Natsika, Josefine Pernes, Oskar Rydén, Maria Verkhovtseva, and Johannes von Römer. 2024. "V-Dem Organization and Management v14" Varieties of Democracy (V-Dem) Project.



1.8 Country Units

The following table contains all country units (and their year coverage) that are included in the V-Dem Dataset. Some countries are coded prior to independence, and some have gaps in their coding periods. For a more detailed description of the country units and their year coverage please consult the V-Dem *Country Coding Units* document.

Name	ID	Coverage
Afghanistan	36	1789–2023
Albania	12	1912–2023
Algeria	103	1900–2023
Angola	104	1900–2023
Argentina	37	1789–2023
Armenia	105	1990–2023
Australia	67	1789–2023
Austria	144	1789–2023
Azerbaijan	106	1990–2023
Baden	349	1789–1871
Bahrain	146	1900–2023
Bangladesh	24	1971–2023
Barbados	147	1900–2023
Bavaria	350	1789–1871
Belarus	107	1990–2023
Belgium	148	1789–2023
Benin	52	1900–2023
Bhutan	53	1900–2023
Bolivia	25	1825–2023
Bosnia and Herzegovina	150	1992–2023
Botswana	68	1900–2023
Brazil	19	1789–2023
Brunswick	363	1789–1867
Bulgaria	152	1878–2023
Burkina Faso	54	1919–2023
Burma/Myanmar	10	1789–2023
Burundi	69	1916–2023
Cambodia	55	1900–2023
Cameroon	108	1961–2023
Canada	66	1841–2023
Cape Verde	70	1900–2023
Central African Republic	71	1920–2023
Chad	109	1920–2023
Chile	72	1789–2023
China	110	1789–2023
Colombia	15	1789–2023
Comoros	153	1900–2023
Costa Rica	73	1838–2023
Croatia	154	1941–2023
Cuba	155	1789–2023
Cyprus	156	1900–2023
Czechia	157	1918–2023
Democratic Republic of the Congo	111	1900–2023
Denmark	158	1789–2023
Djibouti	113	1900–2023
Dominican Republic	114	1789–2023
Ecuador	75	1830–2023



Name	ID	Coverage
Egypt	13	1789–2023
El Salvador	22	1838–2023
Equatorial Guinea	160	1900–2023
Eritrea	115	1900–2023
Estonia	161	1918–2023
Eswatini	132	1900–2023
Ethiopia	38	1789–2023
Fiji	162	1900–2023
Finland	163	1809–2023
France	76	1789–2023
Gabon	116	1910–2023
Georgia	118	1990–2023
German Democratic Republic	137	1949–1990
Germany	77	1789–2023
Ghana	7	1902–2023
Greece	164	1822–2023
Guatemala	78	1789–2023
Guinea	63	1900–2023
Guinea-Bissau	119	1900–2023
Guyana	166	1900–2023
Haiti	26	1789–2023
Hamburg	362	1789–1867
Hanover	357	1789–1866
Hesse-Darmstadt	359	1789–1866
Hesse-Kassel	358	1789–1866
Honduras	27	1838–2023
Hong Kong	167	1900–2023
Hungary	210	1789–2023
Iceland	168	1900–2023
India	39	1789–2023
Indonesia	56	1800–2023
Iran	79	1789–2023
Iraq	80	1920–2023
Ireland	81	1919–2023
Israel	169	1948–2023
Italy	82	1861–2023
Ivory Coast	64	1900–2023
Jamaica	120	1900–2023
Japan	9	1789–2023
Jordan	83	1922–2023
Kazakhstan	121	1990–2023
Kenya	40	1900–2023
Kosovo	43	1999–2023
Kuwait	171	1789–2023
Kyrgyzstan	122	1990–2023
Laos	123	1900–2023
Latvia	84	1920–2023



Name	ID	Coverage
Lebanon	44	1918–2023
Lesotho	85	1900–2023
Liberia	86	1821–2023
Libya	124	1789–2023
Lithuania	173	1918–2023
Luxembourg	174	1815–2023
Madagascar	125	1817–2023
Malawi	87	1900–2023
Malaysia	177	1900–2023
Maldives	88	1900–2023
Mali	28	1900–2023
Malta	178	1900–2023
Mauritania	65	1904–2023
Mauritius	180	1900–2023
Mecklenburg-Schwerin	360	1789–1867
Mexico	3	1789–2023
Modena	351	1789–1859
Moldova	126	1990–2023
Mongolia	89	1911–2023
Montenegro	183	1789–2023
Morocco	90	1789–2023
Mozambique	57	1900–2023
Namibia	127	1900–2023
Nassau	366	1806–1866
Nepal	58	1789–2023
Netherlands	91	1789–2023
New Zealand	185	1841–2023
Nicaragua	59	1838–2023
Niger	60	1922–2023
Nigeria	45	1914–2023
North Korea	41	1945–2023
North Macedonia	176	1991–2023
Norway	186	1789–2023
Oldenburg	364	1789–1867
Oman	187	1789–2023
Pakistan	29	1947–2023
Palestine/British Mandate	209	1918–1948
Palestine/Gaza	138	1948–2023
Palestine/West Bank	128	1948–2023
Panama	92	1903–2023
Papal States	361	1789–1870
Papua New Guinea	93	1900–2023
Paraguay	189	1811–2023
Parma	352	1789–1859
Peru	30	1789–2023
Philippines	46	1900–2023
Piedmont-Sardinia	373	1789–1861
Poland	17	1789–2023
Portugal	21	1789–2023
Qatar	94	1900–2023
Republic of the Congo	112	1903–2023
Republic of Vietnam	35	1802–1975
Romania	190	1789–2023
Russia	11	1789–2023
Rwanda	129	1916–2023



Name	ID	Coverage
Sao Tome and Principe	196	1900–2023
Saudi Arabia	197	1789–2023
Saxe-Weimar-Eisenach	365	1809–1867
Saxony	353	1789–1867
Senegal	31	1904–2023
Serbia	198	1804–2023
Seychelles	199	1903–2023
Sierra Leone	95	1900–2023
Singapore	200	1867–2023
Slovakia	201	1939–2023
Slovenia	202	1989–2023
Solomon Islands	203	1900–2023
Somalia	130	1900–2023
Somaliland	139	1900–2023
South Africa	8	1900–2023
South Korea	42	1789–2023
South Sudan	32	2011–2023
South Yemen	23	1900–1990
Spain	96	1789–2023
Sri Lanka	131	1900–2023
Sudan	33	1900–2023
Suriname	4	1900–2023
Sweden	5	1789–2023
Switzerland	6	1789–2023
Syria	97	1918–2023
Taiwan	48	1900–2023
Tajikistan	133	1990–2023
Tanzania	47	1914–2023
Thailand	49	1789–2023
The Gambia	117	1900–2023
Timor-Leste	74	1900–2023
Togo	134	1916–2023
Trinidad and Tobago	135	1900–2023
Tunisia	98	1789–2023
Türkiye	99	1789–2023
Turkmenistan	136	1990–2023
Tuscany	354	1789–1861
Two Sicilies	356	1789–1860
Uganda	50	1900–2023
Ukraine	100	1990–2023
United Arab Emirates	207	1971–2023
United Kingdom	101	1789–2023
United States of America	20	1789–2023
Uruguay	102	1825–2023
Uzbekistan	140	1789–2023
Vanuatu	206	1906–2023
Venezuela	51	1789–2023
Vietnam	34	1945–2023
Württemberg	355	1789–1871
Yemen	14	1789–2023
Zambia	61	1911–2023
Zanzibar	236	1856–2023
Zimbabwe	62	1900–2023
.	.	.
Total number of countries	202	



1.9 Identifier Variables in the V-Dem and DSP Datasets

1.9.1 Country Name (`country_name`)

Name of coded country. A V-Dem country is a political unit enjoying at least some degree of functional and/or formal sovereignty. For more details on country units consult the V-Dem *Country Coding Units* document.

Response: Text.

1.9.2 V-Dem Country ID (`country_id`)

Unique country ID designated for each country. A list of countries and their corresponding IDs used in the V-Dem dataset can be found in the country table in the codebook, as well as in the V-Dem *Country Coding Units* document.

Response: Numeric.

1.9.3 Country Name Abbreviation (`country_text_id`)

Abbreviated country names.

Response: Text.

1.9.4 Year (`year`)

V-Dem year coded annually from 2000-2023. This variable is included in the V-Dem Country Year as well as Country Date datasets.

Response: Date.

1.9.5 Historical Date (`historical_date`)

This variable is included in the V-Dem Country Date dataset. As of v13 (V-Dem) or v5 (DSP), December 31st, as in 2022-12-31, is no longer the default date. When experts provide ratings through the online interface, the default date is still December 31st. However, in our data processing we add January 1st, as in 2022-01-01, to all variables except election-date specific ones (only present in the V-Dem dataset). This has the effect that December 31st is no longer representative of the full year. This variable also exists in the country-year dataset, but has no direct meaning in that context and should not be preferred over the year column.

Response: Date.

1.9.6 Start of Coding Period (`codingstart`)

The DSP country coding starts in 2000, or from when a country first enjoyed at least some degree of functional and/or formal sovereignty. For detailed information, please see the V-Dem *Country Coding Units* document.

Response: Date.

1.9.7 Gap in Coding Period Starts (`gapstart`)

Time periods when a country does not fulfill V-Dem's coding period criteria are not coded. The date that indicates the gap start is the last date coded before the gap. For more details about V-Dem country coding periods, please see the V-Dem *Country Coding Units* document.

Response: Date.

1.9.8 Gap in Coding Period Ends (`gapend`)

The periods of when a country does not fulfill V-Dem's coding period macriteria are not coded. The date that indicates the gap end is the first date coded after the gap. For more details about V-Dem



country coding periods, please see the V-Dem *Country Coding Units* document.

Response: Date.

1.9.9 Gap index (gap_index)

An index for each country and continuous non-gap, i.e. it is reasonable for interpolation to interpolate only within the same gap index. For more details about V-Dem country coding periods, please see the V-Dem *Country Coding Units* document.

Response: Numeric.

1.9.10 End of Coding Period (codingend)

The DSP country coding ends in 2023, or from when a country formally stopped enjoying at least some degree of functional and/or formal sovereignty. For detailed information, please see the V-Dem *Country Coding Units* document.

Response: Date.

1.9.11 COW Code (COWcode)

Correlates of War (COW) project country codes.

Response: Numeric.

Citation: Correlates of War Project (2017).



2 Digital Society Survey

The Digital Society Survey, designed by the Digital Society Project, contains questions pertaining to the political environment of the internet and social media. The data collected through expert-coded surveys provides information on topics related to coordinated information operations, digital media freedom, online media polarization, social cleavages as well as state internet regulation capacity and approach.

Principal investigators for the Digital Society Project are Valeriya Mechkova, Daniel Pemstein, Brigitte Seim, Steven Wilson.

For more information, please visit www.digitalsocietyproject.org.

Instructions to the coders (as shown in the surveys)

Digital society: The following survey contains questions pertaining to the political environment of the Internet and social media. Please bear in mind the following definitions as you respond to questions on this survey:

The government and its agents include official government organs, such as bureaucracies, courts, intelligence services, and the military, but also unofficial agents, such as officially unaffiliated cyber-warfare operatives who perform services, even “off-book” work, on behalf of the government.

Major political parties include the group of political parties that hold a significant number of seats in national legislative body(-ies), or earn a significant number of votes in elections for the executive. When we ask you to consider “major political parties,” you do not need to consider parties that run in elections but receive only a small minority of seats or votes, or those that receive no seats at all.

We define the Internet as all information that people access over public and private digital networks, worldwide. The Internet includes both publicly accessible digital spaces and private or gated information transmission platforms. The Internet does not include traditional media transmission mechanisms such as paper, television, traditional voice telephone, and radio.

Social media are a subset of Internet platforms that enable normal individuals to create and share content with networks of other people. Social media platforms are available to the public, although content on such networks may be shared privately within subgroups of users. Social media includes both publicly visible, or semi-public platforms, like Facebook, Flickr, Friendster, Google+, Instagram, Myspace, LinkedIn, Twitter, VKontakte, and Weibo and private social networking and messaging platforms like Signal, Slack, Snapchat, or WhatsApp.

Domestic online media is any media source originating in the country in question. For example, the New York Times’ website is domestic online media in the United States, but not in India, even though it operates bureaus in India. Media includes any source reporting on current events or political issues, ranging from well-established brands to newsletters and websites run by an individual.

Cyber security threats include penetration of private digital networks, using means ranging from exploiting software vulnerabilities, password cracking, or social engineering (e.g., tricking individuals into revealing passwords or other information necessary to break into a digital system) to obtain information or disrupt an organization or individual’s use of digital networks and tools. They also include unauthorized alterations of an individual or organization’s digital presence, such as defacing websites and commandeering social media accounts. These threats range from unsophisticated (e.g., exploitation of failure to password protect private networks or use of common passwords by authorized users, and spear phishing) to moderate (e.g., embedding malicious code in emails or exploiting well-known software flaws that organizations have failed to patch), to sophisticated (e.g., exploiting unknown exploits in commonly used software or even embedding exploits into commercial systems unbeknownst to their creators).



Clarification: When we discuss shutting down online content, please consider instances where a website (or websites) have been taken entirely offline as well as instances where a website (or websites) have been slowed down or had access similarly intentionally inhibited, such that use of this website is challenging. In other words, both outright shutting down and more subtle measures that inhibit access should be considered when answering these questions.

Clarification: When we discuss “censorship” or “censoring” content online, we are not concerned with censorship of topics such as child pornography, highly classified information such as military or intelligence secrets, or defamatory speech, unless this sort of censorship is used as a pretext for censoring political information or opinions.

2.1 Coordinated Information Operations

2.1.1 Government dissemination of false information domestic (C) (v2smgovdom)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do the government and its agents use social media to disseminate misleading viewpoints or false information to influence its own population?

Responses:

- 0: Extremely often. The government disseminates false information on all key political issues.
- 1: Often. The government disseminates false information on many key political issues.
- 2: About half the time. The government disseminates false information on some key political issues, but not others.
- 3: Rarely. The government disseminates false information on only a few key political issues.
- 4: Never, or almost never. The government never disseminates false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.1.2 Government dissemination of false information abroad (C) (v2smgovab)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do the government and its agents use social media to disseminate misleading viewpoints or false information to influence citizens of other countries abroad?

Responses:

- 0: Extremely often. The government disseminates false information on all key political issues.
- 1: Often. The government disseminates false information on many key political issues.
- 2: About half the time. The government disseminates false information on some key political issues, but not others.
- 3: Rarely. The government disseminates false information on only a few key political issues.
- 4: Never, or almost never. The government never disseminates false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).



Years: 2000-2023

2.1.3 Party dissemination of false information domestic (C) (v2smpardom)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do major political parties and candidates for office use social media to disseminate misleading viewpoints or false information to influence their own population?

Responses:

0: Extremely often. Major political parties and candidates disseminate false information on all key political issues.

1: Often. Major political parties and candidates disseminate false information on many key political issues.

2: About half the time. Major political parties and candidates disseminate false information on some key political issues, but not others.

3: Rarely. Major political parties and candidates disseminate false information on only a few key political issues.

4: Never, or almost never. Major political parties and candidates never disseminate false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.1.4 Party dissemination of false information abroad (C) (v2smparab)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do major political parties and candidates for office use social media to disseminate misleading viewpoints or false information to influence citizens of other countries abroad?

Responses:

0: Extremely often. Major political parties and candidates disseminate false information on all key political issues.

1: Often. Major political parties and candidates disseminate false information on many key political issues.

2: About half the time. Major political parties and candidates disseminate false information on some key political issues, but not others.

3: Rarely. Major political parties and candidates disseminate false information on only a few key political issues.

4: Never, or almost never. Major political parties and candidates never disseminate false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023



2.1.5 Foreign governments dissemination of false information (C) (v2smfordom)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How routinely do foreign governments and their agents use social media to disseminate misleading viewpoints or false information to influence domestic politics in this country?

Responses:

0: Extremely often. Foreign governments disseminate false information on all key political issues.

1: Often. Foreign governments disseminate false information on many key political issues.

2: About half the time. Foreign governments disseminate false information on some key political issues, but not others.

3: Rarely. Foreign governments disseminate false information on only a few key political issues.

4: Never, or almost never. Foreign governments never disseminate false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.1.6 Foreign governments ads (C) (v2smforads)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How routinely do foreign governments and their agents use paid advertisements on social media in order to disseminate misleading viewpoints or false information to influence domestic politics in this country?

Responses:

0: Extremely often. Foreign governments disseminate false information on all key political issues.

1: Often. Foreign governments disseminate false information on many key political issues.

2: About half the time. Foreign governments disseminate false information on some key political issues, but not others.

3: Rarely. Foreign governments disseminate false information on only a few key political issues.

4: Never, or almost never. Foreign governments never disseminate false information on key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

Convergence: Model parameters with convergence issues: country-date latent trait estimates, universal thresholds, expert reliability, expert threshold, main-country-coded thresholds.



2.2 Digital Media Freedom

2.2.1 Government Internet filtering capacity (C) (v2smgovfilcap)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Independent of whether it actually does so in practice, does the government have the technical capacity to censor information (text, audio, images, or video) on the Internet by filtering (blocking access to certain websites) if it decided to?

Responses:

- 0: The government lacks any capacity to block access to any sites on the Internet.
- 1: The government has limited capacity to block access to a few sites on the Internet.
- 2: The government has adequate capacity to block access to most, but not all, specific sites on the Internet if it wanted to.
- 3: The government has the capacity to block access to any sites on the Internet if it wanted to.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.2 Government Internet filtering in practice (C) (v2smgovfilprc)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How frequently does the government censor political information (text, audio, images, or video) on the Internet by filtering (blocking access to certain websites)?

Responses:

- 0: Extremely often. It is a regular practice for the government to remove political content, except to sites that are pro-government.
- 1: Often. The government commonly removes online political content, except sites that are pro-government.
- 2: Sometimes. The government successfully removes about half of the critical online political content.
- 3: Rarely. There have been only a few occasions on which the government removed political content.
- 4: Never, or almost never. The government allows Internet access that is unrestricted, with the exceptions mentioned in the clarifications section.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.3 Government Internet shut down capacity (C) (v2smgovshutcap)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Independent of whether it actually does so in practice, does the government have the technical capacity to actively shut down domestic access to the Internet if it decided to?

Clarification: A domestic Internet connection is any connection originating physically within the



country, whether over wired, wireless, or satellite networks. This question asks what proportion of potential Internet connections of domestic origin the government has the capacity to render inoperable.

Responses:

- 0: The government lacks the capacity to shut down any domestic Internet connections.
- 1: The government has the capacity to shut down roughly a quarter of domestic access to the Internet.
- 2: The government has the capacity to shut down roughly half of domestic access to the Internet.
- 3: The government has the capacity to shut down roughly three quarters of domestic access to the Internet.
- 4: The government has the capacity to shut down all, or almost all, domestic access to the Internet.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.4 Government Internet shut down in practice (C) (v2smgovshut)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often does the government shut down domestic access to the Internet?

Responses:

- 0: Extremely often. It is a regular practice for the government to shut down domestic access to the Internet.
- 1: Often. The government shut down domestic access to the Internet numerous times this year.
- 2: Sometimes. The government shut down domestic access to the Internet several times this year.
- 3: Rarely but there have been a few occasions throughout the year when the government shut down domestic access to Internet.
- 4: Never, or almost never. The government does not typically interfere with the domestic access to the Internet.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.5 Government social media shut down in practice (C) (v2smgovsm)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often does the government shut down access to social media platforms?

Responses:

- 0: Extremely often. It is a regular practice for the government to shut down access to social media.
- 1: Often. The government shuts down access to social media numerous times this year.
- 2: Sometimes. The government shuts down access to social media several times this year.



3: Rarely. There have been a few occasions throughout the year when the government shuts down access to social media.

4: Never, or almost never. The government does not interfere with the access to social media, except in the cases mentioned in the clarifications section.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.6 Government social media alternatives (C) (v2smgovsmalt)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How prevalent is the usage of social media platforms that are wholly controlled by either the government or its agents in this country?

Responses:

0: Essentially all social media usage takes place on platforms controlled by the state.

1: Most usage of social media is on state-controlled platforms, although some groups use non-state-controlled alternatives.

2: There is significant usage of both state-controlled and non-state-controlled social media platforms.

3: While some state-controlled social media platforms exist, their usage only represents a small share of social media usage in the country.

4: Practically no one uses state-controlled social media platforms.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

Convergence: Model parameters with convergence issues: country-date latent trait estimates, universal thresholds, expert reliability, expert threshold, main-country-coded thresholds.

2.2.7 Government social media monitoring (C) (v2smgovsmmon)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How comprehensive is the surveillance of political content in social media by the government or its agents?

Responses:

0: Extremely comprehensive. The government surveils virtually all content on social media.

1: Mostly comprehensive. The government surveils most content on social media, with comprehensive monitoring of most key political issues.

2: Somewhat comprehensive. The government does not universally surveil social media but can be expected to surveil key political issues about half the time.

3: Limited. The government only surveils political content on social media on a limited basis.

4: Not at all, or almost not at all. The government does not surveil political content on social media, with the exceptions mentioned in the clarifications section.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.



Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.8 Government social media censorship in practice (C) (v2smgovsmcncprc)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: To what degree does the government censor political content (i.e., deleting or filtering specific posts for political reasons) on social media in practice?

Responses:

- 0: The government simply blocks all social media platforms.
- 1: The government successfully censors all social media with political content.
- 2: The government successfully censors a significant portion of political content on social media, though not all of it.
- 3: The government only censors social media with political content that deals with especially sensitive issues.
- 4: The government does not censor political social media content, with the exceptions mentioned in the clarifications section.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.2.9 Government cyber security capacity (C) (v2smgovcapsec)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Does the government have sufficiently technologically skilled staff and resources to mitigate harm from cyber-security threats?

Responses:

- 0: No. The government does not have the capacity to counter even unsophisticated cyber security threats.
- 1: Not really. The government has the resources to combat only unsophisticated cyber attacks.
- 2: Somewhat. The government has the resources to combat moderately sophisticated cyber attacks.
- 3: Mostly. The government has the resources to combat most sophisticated cyber attacks.
- 4: Yes. The government has the resources to combat sophisticated cyber attacks, even those launched by highly skilled actors.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023



2.2.10 Political parties cyber security capacity (C) (v2smpolcap)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Do the major political parties have sufficiently technologically skilled staff and resources to mitigate harm from cyber security threats?

Responses:

0: No. The government does not have the capacity to counter even unsophisticated cyber security threats.

1: Not really. The government has the resources to combat only unsophisticated cyber attacks.

2: Somewhat. The government has the resources to combat moderately sophisticated cyber attacks.

3: Mostly. The government has the resources to combat most sophisticated cyber attacks.

4: Yes. The government has the resources to combat sophisticated cyber attacks, even those launched by highly skilled actors.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3 State Internet Regulation Capacity and Approach

2.3.1 Internet legal regulation content (C) (v2smregcon)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: What type of content is covered in the legal framework to regulate Internet?

Responses:

0: The state can remove any content at will.

1: The state can remove most content, and the law protects speech in only specific, and politically uncontroversial contexts.

2: The legal framework is ambiguous. The state can remove some politically sensitive content, while other is protected by law.

3: The law protects most political speech, but the state can remove especially politically controversial content.

4: The law protects political speech, and the state can only remove content if it violates well-established legal criteria.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.2 Privacy protection by law exists (C) (v2smprivex)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Does a legal framework to protect Internet users' privacy and their data exist?

Responses:

0: No. (Skip to v2smregcap)

1: Yes



Ordering: if 0 no, Skip to v2smregcap

Scale: yes/no

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.3 Privacy protection by law content (C) (v2smprivcon)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: What does the legal framework to protect Internet users' privacy and their data stipulate?

Responses:

0: The legal framework explicitly allows the government to access any type of personal data on the Internet.

1: The legal framework explicitly allows the government to access most types of personal data on the Internet.

2: The legal framework explicitly allows the government to access many types of personal data on the Internet.

3: The legal framework explicitly allows the government to access only a few types of personal information on the Internet.

4: The legal framework explicitly allows the government to access personal information on the Internet only in extraordinary circumstances.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.4 Government capacity to regulate online content (C) (v2smregcap)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Does the government have sufficient staff and resources to regulate Internet content in accordance with existing law?

Responses:

0: No, almost all online activity happens outside of reach of the state, where it lacks the capacity to remove illegal content.

1: Not really. The state has extremely limited resources to regulate online content.

2: Somewhat. The state has the capacity to regulate only some online content or some portions of the law.

3: Mostly. The state has robust capacity to regulate online content, though not enough to regulate all content and all portions of the law.

4: Yes, the government has sufficient capacity to regulate all online content.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.*



(2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.5 Government online content regulation approach (C) (v2smregapp)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Does the government use its own resources and institutions to monitor and regulate online content or does it distribute this regulatory burden to private actors such as Internet service providers?

Responses:

- 0: All online content monitoring and regulation is done by the state.
- 1: Most online content monitoring and regulation is done by the state, though the state involves private actors in a limited way.
- 2: Some online content monitoring and regulation is done by the state, but the state also involves private actors in monitoring and regulation in various ways.
- 3: The state does little online content monitoring and regulation, and entrusts most of the monitoring and regulation to private actors.
- 4: The state off-loads all online content monitoring and regulation to private actors.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see *V-Dem Methodology*).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.6 Defamation protection (C) (v2smlawpr)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Does the legal framework provide protection against defamatory online content, or hate speech?

Responses:

- 0: No. The law provides no protection against Internet defamation and hate speech.
- 1: Not really. The law provides a weak protection and to very limited range of circumstances.
- 2: Somewhat. The law provides some protection against Internet defamation and hate speech but in limited circumstances, or only to particular groups of people.
- 3: Mostly. The law provides protection against Internet defamation and hate speech under many circumstances, and to most groups of people.
- 4: Yes. The law provides comprehensive protection against Internet defamation and hate speech.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see *V-Dem Methodology*).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.3.7 Abuse of defamation and copyright law by elites (C) (v2smdefabu)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr



Question: To what extent do elites abuse the legal system (e.g., defamation and copyright law) to censor political speech online?

Responses:

0: Regularly. Elites abuse the legal system to remove political speech from the Internet as regular practice.

1: Often. Elites commonly abuse the legal system to remove political speech from the Internet.

2: Sometimes. Elites abuse the legal system to remove political speech from the Internet about half the time.

3: Rarely. Elites occasionally abuse the legal system to remove political speech from the Internet.

4: Never, or almost never. Elites do not abuse the legal system to remove political speech from the Internet.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.4 Online Media Polarization

2.4.1 Online media existence (C) (v2smonex)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Do people consume domestic online media?

Responses:

0: Not at all. No one consumes domestic online media. Skip next question if this answer is selected.

1: Limited. Domestic online media consumption is limited.

2: Relatively extensive. Domestic online media consumption is common.

3: Extensive. Almost everyone consumes domestic online media.

Ordering: if 0, skip v2smonper

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.4.2 Online media perspectives (C) (v2smonper)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Do the major domestic online media outlets represent a wide range of political perspectives?

Responses:

0: The major domestic online media outlets represent only the government's perspective.

1: The major domestic online media outlets represent only the perspectives of the government and a government approved, semi-official opposition party.

2: The major domestic online media outlets represent a variety of political perspectives but they systematically ignore at least one political perspective that is important in this society.

3: All perspectives that are important in this society are represented in at least one of the



major domestic online media outlets.

4: All perspectives that are important in this society are represented in many major domestic online media outlets.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.4.3 Online media fractionalization (C) (v2smmefra)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: Do the major domestic online media outlets give a similar presentation of major (political) news?

Responses:

0: No. The major domestic online media outlets give opposing presentation of major events.

1: Not really. The major domestic online media outlets differ greatly in the presentation of major events.

2: Sometimes. The major domestic online media outlets give a similar presentation of major events about half the time.

3: Mostly. The major domestic online media outlets mostly give a similar presentation of major events.

4: Yes. Although there are small differences in representation, the major domestic online media outlets give a similar presentation of major events.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5 Social Cleavages

2.5.1 Online harassment groups (C) (v2smhargr)

Additional versions: *_nr

Question: Which groups are targets of hate speech or harassment in online media?

Clarification: Multiple selection. Choose all that apply.

Responses:

0: Women [v2smhargr_0]

1: LGBTQ groups and individuals [v2smhargr_1]

2: Specific religious groups [v2smhargr_2]

3: Specific ethnic groups [v2smhargr_3]

4: Specific caste [v2smhargr_4]

5: Specific language groups [v2smhargr_5]

6: Specific race [v2smhargr_6]

7: People with physical or cognitive disabilities [v2smhargr_7]

8: People from specific regions [v2smhargr_8]

9: Other (specify in the next question) [v2smhargr_9]

10: No group is a specific target [v2smhargr_10]



Scale: Series of dichotomous scales.

Data release: 9-14.

Cross-coder aggregation: Mean.

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper 2019:1*); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.2 Other online harassment groups (C) (v2smhargrtxt)

Question: Which other groups are targets of hate speech or harassment in online media?

Clarification: Skip if the question does not apply to this country.

Scale: Text.

Data release: 9-14. Available upon request, subject to review and approval.

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper 2019:1*); *V-Dem Codebook* (see suggested citation at the top of this document).

2.5.3 Use of social media to organize offline violence (C) (v2smorgviol)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do people use social media to organize offline violence?

Responses:

0: Frequently. There are numerous cases in which people have used social media to organize offline violence.

1: Sometimes. There are a few cases in which people have used social media to organize offline violence.

2: Never. People have never used social media to organize offline violence.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see *V-Dem Methodology*).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper 2019:1*); Pemstein *et al.* (2024, *V-Dem Working Paper Series 2024:21*); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.4 Average people's use of social media to organize offline action (C) (v2smorgavgact)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do average people use social media to organize offline political action of any kind?

Responses:

0: Never or almost never. Average people have almost never used social media to organize offline political action.

1: Rarely. Average people do not typically use social media to organize offline political action.

2: Sometimes. There are a few cases in which average people have used social media to organize offline political action.

3: Often. There have been several cases in which average people have used social media to organize offline political action.

4: Regularly. There are numerous cases in which average people have used social media to organize offline political action.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.



Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.5 Elites' use of social media to organize offline action (C) (v2smorgelitact)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do domestic elites use social media to organize offline political action of any kind?

Responses:

0: Never or almost never. Elites have almost never used social media to organize offline political action.

1: Rarely. Elites do not typically use social media to organize offline political action.

2: Sometimes. There are a few cases in which elites have used social media to organize offline political action.

3: Often. There have been several cases in which elites have used social media to organize offline political action.

4: Regularly. There are numerous cases in which elites have used social media to organize offline political action.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.6 Types of organization through social media (C) (v2smorgtypes)

Additional versions: *_nr

Question: What types of offline political action are most commonly mobilized on social media?

Clarification: Multiple selection. Choose all that apply.

Responses:

0: Petition signing [v2smorgtypes_0]

1: Voter turnout [v2smorgtypes_1]

2: Street protests [v2smorgtypes_2]

3: Strikes/labor actions [v2smorgtypes_3]

4: Riots [v2smorgtypes_4]

5: Organized rebellion [v2smorgtypes_5]

6: Vigilante Justice (e.g., mob lynching, stalking harassment) [v2smorgtypes_6]

7: Terrorism [v2smorgtypes_7]

8: Ethnic cleansing/genocide [v2smorgtypes_8]

9: Other (specify in the next question) [v2smorgtypes_9]

Scale: Series of dichotomous scales.

Data release: 9-14.

Cross-coder aggregation: Mean.

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023



2.5.7 Other types of organization through social media (C) (v2smorgtypestxt)

Question: What other types of offline political action are most commonly mobilized on social media?

Clarification: Skip if the question does not apply to this country.

Scale: Text.

Data release: 9-14. Available upon request, subject to review and approval.

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); *V-Dem Codebook* (see suggested citation at the top of this document).

2.5.8 Party/candidate use of social media in campaigns (C) (v2smcamp)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: To what extent do major political parties and candidates use social media during electoral campaigns to communicate with constituents?

Responses:

0: None. Major political parties and candidates do not use social media during electoral campaigns to communicate with constituents.

1: A little. Major political parties and candidates rarely use social media during electoral campaigns to communicate with constituents.

2: Somewhat. Major political parties and candidates sometimes use social media during electoral campaigns to communicate with constituents.

3: Substantial. Major political parties and candidates frequently use social media during electoral campaigns to communicate with constituents.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see *V-Dem Methodology*).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.9 Arrests for political content (C) (v2smarrest)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: If a citizen posts political content online that would run counter to the government and its policies, what is the likelihood that citizen is arrested?

Responses:

0: Extremely likely.

1: Likely.

2: Unlikely.

3: Extremely unlikely.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see *V-Dem Methodology*).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023



2.5.10 Polarization of society (C) (v2smpolsoc)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How would you characterize the differences of opinions on major political issues in this society?

Clarification: While plurality of views exists in all societies, we are interested in knowing the extent to which these differences in opinions result in major clashes of views and polarization or, alternatively, whether there is general agreement on the general direction this society should develop.

Responses:

0: Serious polarization. There are serious differences in opinions in society on almost all key political issues, which result in major clashes of views.

1: Moderate polarization. There are differences in opinions in society on many key political issues, which result in moderate clashes of views.

2: Medium polarization. Differences in opinions are noticeable on about half of the key political issues, resulting in some clashes of views.

3: Limited polarization. There are differences in opinions on only a few key political issues, resulting in few clashes of views.

4: No polarization. There are differences in opinions but there is a general agreement on the direction for key political issues.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023

2.5.11 Political parties hate speech (C) (v2smpolhate)

Additional versions: *_osp, *_ord, *_codelow, *_codehigh, *_sd, *_mean, *_nr

Question: How often do major political parties use hate speech as part of their rhetoric?

Clarification: Hate speech is any speech that is intended to insult, offend, or intimidate members of specific groups, defined by race, religion, sexual orientation, national origin, disability, or similar trait.

Responses:

0: Extremely often.

1: Often.

2: Sometimes.

3: Rarely.

4: Never, or almost never.

Scale: Ordinal, converted to interval by the measurement model.

Data release: 9-14.

Cross-coder aggregation: Bayesian item response theory measurement model (see V-Dem Methodology).

Citation: Mechkova *et al.* (2019, *Digital Society Project Working Paper* 2019:1); Pemstein *et al.* (2024, *V-Dem Working Paper Series* 2024:21); *V-Dem Codebook* (see suggested citation at the top of this document).

Years: 2000-2023



3 Appendix: Glossary

Attributes: This section includes the most specific conceptual building blocks we use to discuss democracy and related concepts. Many of our survey questions attempt to ask about a single attribute, for example, "What percentage of the lower (or unicameral) chamber of the legislature is directly elected in popular elections?" Although any of these questions could also be seen as a compendium of multiple attributes (What does it mean to be a legislature? What is a "popular" election?), in a project covering all countries for more than a century, there are degrees of specificity that it is not practical to approach, so attributes are the most specific concepts that we consider feasible to measure.

Country: A sovereign state or semi-sovereign territory. All political units of concern to V-Dem are referred to as countries, even though their status in international law varies — some being colonies others being nation-states or empires.

Cyber Security Threats: Cyber security threats include penetration of private digital networks, using means ranging from exploiting software vulnerabilities, password cracking, or social engineering (*e.g.*, tricking individuals into revealing passwords or other information necessary to break into a digital system) to obtain information or disrupt an organization or individual's use of digital networks and tools. They also include unauthorized alterations of an individual or organization's digital presence, such as defacing websites and commandeering social media accounts. These threats range from unsophisticated (*e.g.*, exploitation of failure to password protect private networks or use of common passwords by authorized users, and spear phishing) to moderate (*e.g.*, embedding malicious code in emails or exploiting well-known software flaws that organizations have failed to patch), to sophisticated (*e.g.*, exploiting unknown exploits in commonly used software or even embedding exploits into commercial systems unbeknownst to their creators).

Domestic Online Media: Domestic online media is any media source originating in the country in question. For example, the New York Times' website is domestic online media in the United States, but not in India, even though it operates bureaus in India. Media includes any source reporting on current events or political issues, ranging from well-established brands to newsletters and websites run by an individual.

Geographic Group: Geographic group refers to those living in rural or urban areas. Urban areas are defined as an area that meets the following conditions: population density exceeds a threshold of 150 persons per square kilometer, there is access to a sizeable settlement of 50,000 people or more within some reasonable travel time, for example 60 minutes by road. (World Development Report, 2009: 54).

Government: The executive branch of the government, including its head of state (HOS) and/or head of government (HOG) — whichever is most prominent, or both if they are both powerful — along with the cabinet, ministries, and top civil servants. We are only concerned here with the government that actually resides within the country or semi-sovereign territory. Thus, in a typical British colony the government would include the governor-general and his local administration but not the King/Queen of England or the government of England.

Government and its Agents: The government and its agents include official government organs, such as bureaucracies, courts, intelligence services, and the military, but also unofficial agents, such as officially unaffiliated cyber-warfare operatives who perform services, even "off-book" work, on behalf of the government.

Internet: We define the Internet as all information that people access over public and private digital networks, worldwide. The Internet includes both publicly accessible digital spaces and private or gated information transmission platforms. The Internet does not include traditional media transmission mechanisms such as paper, television, traditional voice telephone, and radio.



Major Political Parties: Major political parties include the group of political parties that hold a significant number of seats in national legislative body(-ies), or earn a significant number of votes in elections for the executive. When we ask you to consider "major political parties", you do not need to consider parties that run in elections but receive only a small minority of seats or votes, or those that receive no seats at all.

National Government: The highest level of aggregation recognized by the V-Dem project. Refers to the national government of a sovereign state or the territorial level of government for a semi-sovereign colony or territory. Thus, the "national" government of India prior to independence — the British Raj — was situated in New Delhi, *not* in London — even though decisions affecting the Indian colony were often made in London.

Political Groups: Political groups are defined as those who are affiliated with a particular political party or candidate, or a group of parties/candidates. A common form of partisan exclusion is when state services or regulations are implemented in a way that seeks to reward incumbent political supporters and punish non-supporters.

Political Party: An organization that nominates candidates for public office. The term includes a longstanding coalition such as the CDU/CSU in Germany if that coalition functions in most respects like a single party. Sometimes, the identity of a party is obscured by name changes. However, if the party changes names but retains key personnel and is still run by and for the same constituencies then it should be considered the same organization.

Public Authorities: Includes the government as well as subnational governments, agencies, parastatals, and the like. Compare State.

Semisovereign Territory: This refers to a country that is not fully sovereign but nonetheless exercises some — at least minimal — level of self-determination. Many of the countries of concern to this project began as colonies of an empire. If a country moved from semi-sovereign status to sovereign status over the course of the twentieth century — maintaining comparable borders — then we want to code both entities. Likewise, we want to include countries like Taiwan that are not universally recognized as sovereign but nonetheless enjoy self-determination (in part or in full).

Most questions pertaining to semi-sovereign territories ask you to reflect on the practices and institutions located within that territory — rather than the empire or nation-state that may claim ultimate sovereignty over the territory. Thus, a question about the government or judicial bodies seated within a British colony would refer to the governor-general and his local administration rather than the King/Queen or government of England.

Social Group: A social group is differentiated within a country by caste, ethnicity, language, race, region, religion, migration status, or some combination thereof. (It does *not* include identities grounded in sexual orientation, gender, or socioeconomic status.) Social group identity is contextually defined and is likely to vary across countries and through time. Social group identities are also likely to cross-cut, so that a given person could be defined in multiple ways, *i.e.*, as part of multiple groups. Nonetheless, at any given point in time there are social groups within a society that are understood — by those residing within that society — to be different, in ways that may be politically relevant. Contrast Identity group.

Social Media: Social media are a subset of Internet platforms that enable normal individuals to create and share content with networks of other people. Social media platforms are available to the public, although content on such networks may be shared privately within subgroups of users. Social media includes both publicly visible, or semi-public platforms, like Facebook, Flickr, Friendster, Google+, Instagram, Myspace, LinkedIn, Twitter, VKontakte, and Weibo and private social networking and messaging platforms like Signal, Slack, Snapchat, or WhatsApp.



State: A political organization that organizes compulsory domination over a fixed territory on a continual basis.

Variable: A measure of a small number of attributes. Synonymous with "indicator."