Cautionary Notes

V-Dem is firmly committed to full transparency and release of the data we have. 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 starting from 1900, originally coded to 2012. With the updates covering 2013–2020 it has for a few country–variable combinations been impossible to achieve that target. We have found that this at times result in significant changes in point estimates as a consequence of self-selected attrition of Country Experts, rather than actual changes in the country. We therefore strongly advise against using point estimates for country-variable-years with three or fewer (≤ 3) ratings. We suggest to filter these 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 from v7 and onwards.

- Point estimates can jump around slightly due to the simulation-based nature of the estimation process and expert turnover. Consumers of the data should therefore always be attentive to the uncertainty about the estimates, which provides vital information about the degree to which one can be certain that a change in scores reflects an actual change in the level of the concept being measured.

- We constantly improve the coding of factual data (A) to make it as accurate as possible. This may result in changes at the index-level.

- Observations for *Exclusion* and *Legitimation* indicators (section 3.13 and section 3.14) with less than 3 coders per country-date (*_nr < 3) have been removed. Furthermore, observations for *Exclusion indices* (section 5.6) have been removed if not at least 3 components have at least 3 coders per country-date.

- Observations for *Civic and Academic Space* indicators (section 3.15) with less than 3 coders per country-date (*_nr < 3) have been removed. Furthermore, observations for the *Academic Freedom Index* (section 5.15.1) have been removed if not at least 3 components have at least 3 coders per country-date.

- These variables had issues with convergence: v2caconmob, v2capolit, v2clgeocl, v2elboycot, v2elintim, v2elpaidig, v2elpdcamp, v2elpeace, v2exremhsp, v2juaccnt, v2jucomp, v2lginvstp, v2lgotovst, v2peasbgen, v2peasjgeo, v2smmefra, v2x_neopat, v2x_rule, v2xlg_legcon, v2xnp_pres, v2xpe_exlgender, v2xpe_exlgeo. Please see individual codebook entries for additional information. For details on interpreting convergence information, please refer to 1.4.5, the Methodology Document and Pemstein et al. (2021).

- We further ask you to use the following percentage % variables with caution:
  - Female journalists (v2mefemjrn)
  - Weaker civil liberties population (v2clsnlpct)

- *Historical V-Dem:* In the coding of several Historical V-Dem A type variables, the historical part of the time series — including 20 years of overlap with the "contemporary" time series (typically 1900–1920) — were conducted completely independently from the existing coding in the original V-Dem dataset, by one or more new coders. For many of these historical variables, we have gone through and checked the consistency of the coding, further scrutinized the sources, and determined which coding represents the most appropriate score after deliberation. We have subsequently made the appropriate adjustments to the data.

  For other historical A variables we have yet to finalize this process. For these variables, the scores reported for the overlap period (typically 1900–1920) in the dataset are the "contemporary" V-Dem scores, by default. This means that for some countries, where there is disagreement in the historical and contemporary coding in the starting year for the contemporary time series (typically 1900), there may be artificial changes between that year and the year before that do not necessarily reflect a real-world change in the political system in the country. Hence, we

advise users to exert caution before running analysis on the entire time series extending across both the historical and contemporary coding periods.

Please also note that for the variables where there is not full correspondence between the historical (1789–1920) and contemporary (1900–2020) coding, the historical coding of the variables is also provided in their original form as separate variables, carrying a "v3" rather than a "v2" prefix on the variable tag. These "v3" variables are gathered together with a number of new (A and C type) variables that are currently only coded for the Historical V-Dem sample, in a separate section of the codebook.