Data management plan

A Data description

The project draws on several existing sources of data relating to "ground truths". We list these data sources below, together with an indication of their coverage. Data sources with partial coverage, or which are not regularly updated, will be supplemented by the project team.

- electoral democracy, data from Coppedge et al. (2016) (regularly updated, complete geographic coverage)
- for proportionality, data from Döring and Manow (2018) and Bormann and Golder (2022) (regularly updated, partial geographic coverage)
- for congruence, data from Düpont et al. (2022) (regularly updated, complete geographic coverage, partial party coverage)
- for accountability, data from Döring and Manow (2018) (regularly updated, partial geographic coverage)
- for descriptive representation, data on age- and gender-representation by Stockemer and Sundström (2022) (regularly updated, complete geographic coverage)
- for promissory representation, data from Thomson et al. (2017) (not updated, partial geographic coverage)

The project will generate new data in the form of public opinion surveys. This data will be commissioned from survey providers, and will be initially provided in different formats. Where providers supply information which is not relevant to the study aims and objectives, but which could be used to identify individuals, such data will either be removed (detailed geographic location; IP range) or coarsened (age groups) to prevent identification. Data shared between national teams will be regarded as "anonymous" data. The five different national surveys will be combined into a single master file. This master file will be tabular data, with each row corresponding to a row, and each column corresponding to a respondent characteristic or a question answer. This tabular data will be stored as a comma separated values (CSV) file.

The project will additionally generate additional data in the form of interview notes and videos relating to the work package on experiential arguments. Interviews with survey participants who expressed a willingness to be recontacted and interviewed about their support

for an alternative electoral system or form of government will be regarded as "working materials" only: that is, they shall be anonymous, may (according to the interviewee's preference) be either digitally recorded using an audio recorder or recorded in note form by interviewers, and will be destroyed after the production of the video used as the basis for the survey experiment. The video produced as part of this work package shall be included with the replication materials for any corresponding academic output.

Finally, the project will generate several incidental data files in the course of pursuing academic dissemination. This includes R source code and graphs. The nature and number of these additional data items will depend on the research and publication process. These incidental data files are covered in the section "Preservation and sharing after project ends".

B Ethical and legal aspects

We rely on public opinion surveys which ask respondents about their current or former migration status, and their political beliefs. This kind of data is "special category data" under the General Data Protection Regulation (GDPR). The processing of this data is necessary in order to research the research objectives of the project. Because this data is special category data, survey participants must give their explicit consent for their data to be processed. This consent is given to the survey provider, who is the data processor for these purposes. Survey providers release data to their customers after having removed data which could be used to identify the respondent. At this point the data cease to be "personal data" under the scope of the GDPR. However, because data protection regimes vary across the five countries included in the project, we intend to take further steps to protect the anonymity of the respondents, by removing detailed respondent location and IP ranges. In addition to the responses to specific survey questions, we retain information only on:

- the respondent's country of origin (or in the case of former migrants, their previous host country)
- the respondent's age group (age groups to be determined)
- the respondent's electoral district
- the respondent's sex or gender
- the respondent's highest level of educational qualifications
- the party the respondent voted for in the previous national election

These are typically variables included in survey releases for the purposes of survey weighting. No combination of these variables allows individuals to be uniquely identified. At the time of writing, no census provider in the five countries in the study reports information on detailed country of birth at the level of the electoral district. Whilst other census products do cross-tabulate two or more of the variables listed above, these categories are large enough to prevent the identification of particular individuals.

These ethical and legal aspects concern the survey work packages, which form the bulk of the empirical work of the project. There are additional ethical aspects to the use of migrant arguments in survey experiments. These relate to (i) consent of those participating

in follow-up interviews to the use of their arguments and permission for those arguments to be recorded by an actor; and (ii) the perception of these arguments by survey participants. Regarding the first issue: after the interview, we'll prepare scripts and ask interviewees for their consent to use these arguments and (optionally) consent to use their first name. Regarding the second issue: there is a risk that survey participants might believe that the actor is an actual migrant; in order to guard against any risk of (inadvertent) deception, we'll add a chiron at the bottom of the video of the form "Thomas (New Zealand) (words spoken by an actor)".

C Responsibilities and management methodology

HANRETTY will have overall responsibility for the collection of survey data as part of WP1/2. Leaders of national teams will liaise with survey companies in their own countries, and will work with HANRETTY to ensure that the surveys are as comparable as possible, taking into account requirements of translation and local circumstances (i.e., the particular Swiss form of government).

DASSONNEVILLE, GOLDER and GOLDER will have overall responsibility for the collection of "ground truth" data (WP3) which is used in WP4. This responsibility will be delegated a post-doctoral research associate who will lead on the collection of all "ground truth" data, except for the data on electoral democracy, which will be collected by STANLEY. This post-doctoral research associate will in turn work with research assistants recruited by different national teams, who will have particular responsibility for collecting data on promissory representation.

Responsibility for the collection of data as part of WP6 (experiential arguments) will go to a post-doctoral research associate recruited in the UK and supervised by HANRETTY.

Overall responsibility for data documentation and preservation with lie with HANRETTY. HANRETTY will work with team members during and following the article submission process, preparing documentation so that the eventual replication archives satisfy the requirements of peer-reviewed journals and also provide a complete record of the project's work.

D Publication formats, standards, mechanisms and repositories

We describe the data formats, standards and repositories according to the different types of data we generate.

D.1 Survey data

Regardless of the format used by polling providers to transmit the data to the project team, pooled data from the different national surveys will be stored as a comma separated values

(CSV) file. To maximize readability and discoverability, entries in this CSV will be stored as human-readable survey answers. Thus, the answer to question 2 in Box 1 will be stored as "Much better at promise keeping in [home country]", rather than as a numeric code. A limited number of variables will follow standard schema – countries will be recorded as ISO3166-alpha3 codes; and past party vote will be recorded using the PartyFacts schema (https://partyfacts.herokuapp.com).

D.2 Ground truth data

The ground truth data will be stored as a set of seven comma separated values files: one for each meso-level democratic value. Thus, there will be one file for levels of electoral democracy, one file for levels of accountability, and so on. Entries in these files will correspond to observations of countries at different points in time. For some files, time-points will be years. For example: the file for electoral democracy will give values of electoral democracy for the ten years preceding the date of the survey. For other files, time-points will be election years. For example: the file on accountability will give values of accountability at each of the three (first-order) national elections preceding the date of survey. We report the individual values rather than period averages for each country in order to maximize the re-use potential of the data.

A separate file giving details on promissory representation will also be produced. This data file will have "party pledges" as the unit of analysis, such that each row will list a governing party, a specific manifesto pledge, and whether that pledge was (partially) fulfilled or not. In order to maximize the re-use potential of the data, we will mirror the structure of the data set released by Thomson et al. (2017).

D.3 Survey experiment data

The survey experiment data generated by WP6 will be stored as a CSV file, recording the survey arm and responses regarding institutional form. As with the survey data generated under WP1/2, the responses will be stored as human-readable response options rather than as codes.

The video recordings made during WP6 will also be made available so that researchers can view the treatment. The video recordings will be made available as MP4 files using open source codecs.

E Preservation and sharing after project ends

We adopt a twin-track approach to data preservation.

The first track operates through journals' own replication policies. Political science has enthusiastically adopted requirements for data sharing, and the leading journals in the discipline operate their own data repositories – most based on the Dataverse schema. Whilst data preservation efforts under this track will naturally depend on the journals in which our academic outputs appear, this track is the most "natural" track for political scientists, and maximizes the chance that the data will be discoverable, since publishers include a link to replication data as part of the publication process. These replication materials will include the particular R code that we use to generate our findings.

The second track is a "whole project" track, and will see us upload the project materials described above to a Harvard Dataverse repository. The Harvard Dataverse is open to researchers from all universities, not just Harvard, and offers persistent storage for research projects. Harvard Dataverse offers storage in near-perpetuity, and aims to implement FAIR principles.

Data released under the first track will depend on the timing of journal publication decisions, but data will be made available without restrictions on reuse at the time of publication. Data released under this track may be specific to a particular academic output, such that replication data for the proposed academic output "Just noticeable differences in democracy" may be made available without other survey data being made available.

Data released under the second track will be made available as soon as:

- all data collection for the project is complete, *and*
- all academic outputs from WP4 has been published, *or*
- three months after the end of the project, whichever comes sooner

Data will continue to be publicly available without time limit.

We have listed these two tracks for data preservation. Team members will maximize the success of these tracks by including links to replication materials on their personal webpages, but such personal webpages will not be used to store replication data.

References

Bormann, N.-C., & Golder, M. (2022). Democratic electoral systems around the world, 1946–2020. *Electoral Studies*, *78*, 102487.

Coppedge, M., Lindberg, S., Skaaning, S.-E., & Teorell, J. (2016). Measuring high level democratic principles using the V-Dem data. *International Political Science Review*, *37*(5), 580–593.

Döring, H., & Manow, P. (2018). Parlgov: Parties, elections, cabinets. www.parlgov.org

Düpont, N., Kavasoglu, Y. B., Lührmann, A., & Reuter, O. J. (2022). A global perspective on party organizations. validating the varieties of party identity and organization dataset (V-Party). *Electoral Studies*, *75*, 102423.

Stockemer, D., & Sundström, A. (2022). Introducing the Worldwide Age Representation in Parliaments (WARP) data set. *Social Science Quarterly*, *103*(7), 1765–1774.

Thomson, R., Royed, T., Naurin, E., Artés, J., Costello, R., Ennser-Jedenastik, L., Ferguson, M., Kostadinova, P., Moury, C., Pétry, F., et al. (2017). The fulfillment of parties' election pledges: A comparative study on the impact of power sharing. *American Journal of Political Science*, *61*(3), 527–542.