

Data management plan (DMP)

1 Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

Type of data:

Our project will generate individual-level data of voters containing the following information:

- Socio-demographic information (e.g., gender, age, occupation, educational level)
- Information pattern with regard to elections (e.g., internet and news consumption, activities on social media)
- Political profile (e.g., left-right orientation, party preferences, standpoint on specific policy issues)
- Electoral behavior (e.g., political participation, party choice)

Format and volume of the data:

The data will be typical survey data (numerical data) and only very few parts will consist of text data (e.g., from open questions in surveys). Therefore the data will be formatted as UTF-8 CSV-files.

The dataset's size will be rather small (10-20MB).

1.2 How will the data be collected, observed or generated?

Data collection:

The project will conduct a two-wave online survey among voters for the 2019 cantonal elections in the canton of Zurich. The survey sample will be drawn from a file containing all eligible voters from the Swiss Federal Statistical Office. The project requires a rather large number of respondents (5-6,000 after the second wave). Based on experiences of comparable survey designs in previous projects it is to expect that the project will face a large number of non-respondents (for the first wave) and dropouts (for the second wave), therefore it is required to invite an especially large number of voters to participate in the survey (about 30,000).

The data collection will be conducted as follows:

- 7-8 weeks before the election we will invite 30,000 voters to participate in the study and to answer the first wave of the online-survey. We expect to receive 10-12,000 answers.
- 3-4 weeks before the election we will split the respondents in two groups: the study group (80% of first-wave respondents) will receive a "treatment" in form of a letter asking them to use the Voting Advice Application "smartvote" (www.smartvote.ch), which will be provided to them on a dedicated website for this study and so allowing to track the use of smartvote by the respondents (and to link this data to the survey data). The control group will not receive any "treatment".
- 1 week after the election all first-wave respondents will receive the invitation to answer the second wave of the survey.

Data quality:

Data quality will be ensured by applying standard measures (e.g., pre-testing the questionnaires, validity and reliability tests after the data collection).

1.3 What documentation and metadata will you provide with the data?

The data will be provided as UTF-8 formatted CSV-files (readable by all statistical software packages) and accompanied by a text file with metadata and a codebook.

The metadata will contain information on study design, sampling methodology, fieldwork and also information on methods used to construct additional variables based on the raw data and available weights.

The codebook will contain variable-level details, coding and labelling schemes as well as the number of answers and the frequencies of answer options for each question/variable.

Both the metadata file and the codebook will be available as text and as PDF file. They will provide all information necessary for a accurately and effectively secondary analysis of the data.

2 Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

During the data collection the respondents can be identified by the project team. There is no possibility to conduct such large scale quasi-experiments under anonymity for all respondents. Nevertheless, we can ensure the respondents a high level of privacy and protection of their personal data. First, our project will be bound to the comprehending and tight regulations of the Swiss Federal Statistical Office (see <https://www.bfs.admin.ch/bfs/en/home/basics/census/national-census-integrated-system/sampling-frame.html>). Second, only a very small number (not more than four) of persons will have access to the data during the data collection. And third, immediately after concluding the data collection the dataset will be anonymized.

In addition to that, there are no specific ethical, legal or security issues to consider.

2.2 How will data access and security be managed?

We consider the measures to safely store the data during the project as described in point 3.1 as sufficient. The project will use Qualtrics for conducting the online-surveys. As soon as the surveys are closed we will transfer the data to our own servers (see point 3.1) and delete any data on the Qualtrics website. Qualtrics is hosted in the United States, whereas the university servers are hosted in Switzerland.

2.3 How will you handle copyright and Intellectual Property Rights issues?

All data is owned by the University of Lausanne and will be made available under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

3 Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

The projects dataset will have a manageable size of less than 20MB. And the collected data will be completely anonymized. Therefore there is no need for a special IT infrastructure for the duration of the project. The data will be stored on dedicated password protected drives/servers of the involved university departments.

These servers are protected as all other university servers against digital or physical threats (firewalls, automated backups, storage of backups in a second university building).

Finally, members of the project team (2-4 persons) will be allowed to save the data on their own computers/laptops for the duration of the project.

3.2 What is your data preservation plan?

As already mentioned the datasets of this project are neither especially large nor sensitive with regard to their content (due to the complete anonymization). There is also no legal obligation to delete the data or parts of the data. Therefore, it should be sufficient to make the data available at one of the established data repositories (e.g., FORS or Harvard Dataverse; see point 4.1) with regard to the longterm storage of the data after the completion of the project.

The data will be saved as UTF-8 formatted CSV-files. Additional information like metadata and codebooks will be provided as text files (readable by any text editor) and as PDF files.

4 Data sharing and reuse

4.1 How and where will the data be shared?

The complete dataset will be shared on the FORS data repository after completion of the project. We are aware that FORS does not fulfill all requirements of the FAIR Data Principles yet, but it is a realistic option to expect that FORS will fulfill all FAIR requirements in the near future. If against all expectations this will not be the case, the project will publish the dataset in an alternative repository, which fulfills the FAIR requirements.

The datasets used for the projects publications will be made publicly available on the websites of the specific scientific journals, which agreed to publish the projects' articles. In cases, where these journals do not publish the data according to the FAIR Data Principles the project will make them available in a data repository of its choosing (e.g., Harvard Dataverse, <https://dataverse.harvard.edu/>).

4.2 Are there any necessary limitations to protect sensitive data?

All data will be publicly available. There are no additional legal, ethical or confidentiality restrictions.

The all-inclusive dataset will be made publicly available latest at the time of the projects

completion. The specific datasets used for the project's publications will be made publicly available at the time of the publication.

4.3 All digital repositories I will choose are conform to the FAIR Data Principles.

Yes

4.4 I will choose digital repositories maintained by a non-profit organisation.

Yes