Project Number: **890187**

Project Acronym: **PHASE**

Project title: **How is Populism and Health Associated in Europe (PHASE)? A multilevel analysis of the bi-directional interrelationship between populism and ill health**

DATA MANAGEMENT PLAN

1. Data Summary

The purpose of this document is to describe the data management life cycle for all data sets that will be collected, processed, or generated by the PHASE project. The Data Management Plan (DMP) outlines how the research data will be handled during the project and after it is completed. It describes what data will be collected/generated, what methodology and standards are used, whether and how this data will be shared and made open, and how it will be curated and preserved. The Plan describes the measures taken to enable third parties to access, mine, exploit, reproduce, and disseminate (free of charge for any user) the research data. The project's DMP is based on the template provided in the Guidelines on FAIR Data Management in Horizon 2020.

Following the Open Science and H2020 RRI models, participating in the Pilot on Open Research Data in Horizon 2020, the data collected during the project will be released online. This first version of the DMP serves as starting point and guideline for the researchers in the PHASE project. The more elaborated version will be updated in the final stage of the project.

The overarching goal of the PHASE project is to explore the bi-directional association between populism and health. The substantive, research-related work packages of the project are set up as follows:

- WP3: Create a hierarchical panel dataset, nesting individuals in EU regions, and measure the impact of ill health on support for populist parties.
- WP4: Adopt a coding scheme for populists in government and conduct a dynamic multilevel analysis of their impact on health across Europe.
- WP5: Analyze one pivotal policy episode in a selected country using quasi-natural experimental settings.

WP3 analyses how regional-level health and socioeconomic characteristics are associated with individual-level political orientations. Besides, WP3 also includes a systematic review of the impact of health on populism to help to systematize the existing knowledge and aid conceptual development. WP4 analyses the cross-country multilevel evidence on the effect of populists in power (national level) on health inequalities (individual level). Finally, WP5 comprises a case study to identify the structure of populist governance and the role of the policy regime in mitigating or exacerbating the populism – health mechanism.

The data collection relies on re-using already existing data sources, as detailed below. All the data used in the PHASE project is available publicly. The European Social Survey (ESS) provides the backbone of the project. Each wave of the ESS core modules contains information on subjective health, political orientation, and crucial demographic and socioeconomic control variables from 2002 to 2016 (every second year), covering every EU member country, though not every country is included in each wave. The ESS individual survey data can be linked to regional administrative data. Codes for populist political parties are available from the Manifesto Project (Social Science Research Centre, Berlin). International agencies provide data necessary for the case studies, which can be further extended by qualitative data.

The table below summarizes the most important variables, as well as the sources of the data:

WP	Variables	Source of data
WP3	**Individual-level political**: satisfaction with democracy, trust in institutions, confidence in the economy, interest in politics, party affiliation and voting, left-right orientation, libertarian-authoritarian orientation, national sentiment **Regional-level health and socioeconomic**: life expectancy, crude death rates, cause-specific mortality, deaths of despair,	**Individual-level political**: ESS core modules: variables on political orientations, socioeconomic and demographic factors to be used as control variables, covering the 2000-2016 period for every EU country **Regional-level health and socioeconomic**: The Quality of Government EU Regional Dataset,

	population, age structure, migration, dependency ratio; unemployment rate, employment rate by gender and age, industrial employment, GDP per capita, people at risk of poverty or social exclusion, highest level of education of population	University of Gothenburg: The Quality of Government Institute contains time series data at NUTS2 level for the 2000-2014 period. Eurostat Regional Data can be downloaded to update the quality of government dataset
WP4	**Individual-level health**: subjective general health, hampered in daily activities by illness/disability/infirmity/mental problem, trust, alienation (take part in social activities, anyone to discuss personal matters with) **National-level political**: codes for populism covering political parties in the EU.	**Individual-level health**: ESS core modules contain variables on subjective health, social exclusion, as well as socioeconomic and demographic factors that can be used as control variables. **National-level political**: The Manifesto Project (codes for populism) covering political parties in Europe; can be updated from the literature
WP5	**National-level policies**: Government expenditure by function, categorical information on social and health policies **National-level health data**: mortality rates by age, sex and cause (1980-2017) **Qualitative data on policy decisions**: detailed histories of key policymaking events, including actors, reasons, target groups, expected costs and benefits	**National-level spending**: Eurostat COFOG dataset; OECD Social Expenditure Database, which includes standardized information on spending in cash and in-kind, Comparative Welfare Entitlements Dataset **National-level health**: WHO European Health for All Database **Qualitative data**: online media / desk research / secondary literature

During the project, the ESS, Eurostat, WHO, and other datasets will be downloaded, cleaned, and merged. Only those variables will be kept that are relevant for the research questions set out in the PHASE project. The final datasets for the three work packages are expected to be less than 1 gigabyte. The project-related data files will be stored as STATA dta files, excel files, or csv files. STATA also allows for do-files that code all manipulations in the data. All data sets generated in STATA will thus be presented as a collection of raw source files (with reference) and a series of .do-files that allow for exact replication of aggregation, manipulation, and analysis of the data. These .do-files are published with the raw and final cleaned data files. In addition to the quantitative data, desktop data collection will be conducted by monitoring country-level policymaking as documented by international agencies (i.e., IMF) and published in specialized outlets (e.g., covid-19 policy trackers). The qualitative data will be stored in excel and word document files.

Latest at the end of the project, or upon the publication of the relevant articles, all the data will be made publicly available, benefiting the research community by analyzing questions related to populism and health (‘data utility’). Depending on funding availability, articles will be published in ‘gold’ open-access format or ‘green’ open access, i.e., accepted versions deposited online.

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Within the duration of the PHASE project, only the directly responsible researchers (primarily the Fellow and other researchers who might join as co-authors) have access to the data files. They are thereby also responsible for the integrity of the datasets and must carefully document collection and any manipulations made to the data. Data will be made public only after the publication of the reports and deliverables.

All Datasets are Deliverable as a data file and will be labeled with a persistent identifier received upon depositing the dataset. There will be a separate report provided to all datasets, describing in detail the collection and presenting the descriptive statistics and data manipulations of each data series in the dataset; and will be stored alongside the data. Common metadata that apply to all studies in the PHASE project on study level will include name, description, authors, date, subproject, persistent identifier, accompanying publications, source. For such generic metadata, the *Dublin core or DDI metadata standard* (https://ddialliance.org/) will be used. DDI is a widely used international standard for describing data from the social, behavioral, and economic sciences. The project will also re-use existing metadata from the European Social Survey and the various Eurostat data sources. For WP5, a new metadata template must be developed that is suitable to qualitative policy data.

In order to better organize the data and save time, the file naming convention will be used to enable the titling of folders, documents, and records consistently and logically. The data will be available under a filename composed of the project Acronym, sub-project title, dataset name, and version number. For example: PHASE_HealthPopulism_ESS_v1.dta, PHASE_HealthPopulism_Eurostat_v1.dta, etc. Reports will be stored under corresponding names. Furthermore, specific project/data identifiers will be assigned. All variables are given short names, small caps, separating variables at different levels of analysis. A complete codebook will be provided for each dataset, with definitions and descriptive statistics.

Search keywords will include the name of the project (PHASE) and keywords relatable to the project's subject matter, such as populism, health, European Social Survey, multilevel modeling, covid-19, etc.

2.2. Making data openly accessible

The quantitative datasets related to WP 3 and WP4 and the qualitative dataset related to WP5 will be publicly available. The research data and the publications from the PHASE project will be deposited on the SocArXiv (https://osf.io/preprints/socarxiv) platform.

The research data and the publications from the PHASE project will be deposited on the SocArXiv platform. SocArXiv, an open archive of the social sciences, provides a free, nonprofit, open-access platform for social scientists to upload working papers, preprints, and published papers, with the option to link data and code. SocArXiv is dedicated to opening up social science, reaching more people more effectively, improving research, and building the future of scholarly communication. Every paper on SocArXiv is automatically associated with a Project on the Open Science Framework platform. Researchers can attach data, code, or other research materials to their papers by including them in the associated project. SocArXiv encourages sharing data and code on the Open Science Framework in the spirit of openness and transparency.

SocArXiv is a partner of the nonprofit Center for Open Science (COS), and SocArXiv as a web platform is hosted on the OSF Framework. As a legal entity, SocArXiv is an administrative unit at the University of Maryland, College Park (UMD). Grants and donations come to UMD and then are administered by SocArXiv Director Philip Cohen. The SocArXiv Steering Committee is actively involved in an advisory role, and volunteer working groups provide additional assistance. UMD provides fiscal oversight, ensuring that all expenses paid are justified according to University standards. In addition, the whole Center for Open Science system is backed up by "a $250,000 preservation fund for hosted data in the event that COS had to curtail or close its offices. If activated, the preservation fund will preserve and maintain read access to hosted data." Since the development of SocArXiv was first announced in July 2016, researchers have deposited about 5000 papers.

SocArXiv fully supports the use of metadata, including co-authors, DOI, ORCID, etc. Papers, datasets, and project files uploaded to SocArXiv receive a Digital Object Identifier (DOI). In addition, if a paper previously has a DOI from a different publisher, that information can be included when uploading it to SocArXiv, which will allow the paper to link to the latest published version. Note also that the OSF platform on which SocArXiv runs creates a persistent URL for every paper. Google Scholar is also indexing and linking to SocArXiv papers (and other papers on the OSF Preprints server), which greatly facilitates the discoverability of preprint articles and related data files.

The project-related data files will be stored as STATA dta files, excel files, or csv files. The project will also result in STATA do files. The qualitative data will be stored in excel and word document files. Access to data deposited on SocArXiv is free of charge, does not require registration. Since the software used to create and access the data are widely used in the social sciences, there is no need for separate documentation about the software needed to access the data.

The Fellow has already deposited papers on SocArXiv, has experience in its features, and has explored the appropriate arrangements. There will be no restrictions on use. There is no need for a data access committee. The data can be used by anyone anonymously.

In addition, the Fellow will further explore options to deposit project-related data in The Harvard Dataverse Repository (https://dataverse.harvard.edu/). Since 2012, the Institute for Quantitative Social Science has collaborated with Harvard Library and Harvard University Information Technology (HUIT) to make the Harvard Dataverse openly available to researchers and data collectors worldwide from all disciplines to deposit data. In this collaboration, IQSS leads the development of the open-source Dataverse software; and in conjunction with the Open Data Assistance Program at Harvard (a collaboration with Harvard Library, the Office for Scholarly Communication and IQSS), provides user support for the Harvard Dataverse. The Library Technology Services at HUIT provides hosting and backups support of the Harvard Dataverse. The Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data. Each Dataverse collection is a customizable collection of datasets (or a virtual repository) for organizing, managing, and showcasing datasets.

2.3. Making data interoperable

The data produced in the project is fully interoperable as it is packaged in standard dta, excel, csv, word, and pdf files. Anyone with minimum computer skills and access to Microsoft Office, STATA, and pdf reader will be able to open and use the data (append, record, etc.). All of the metadata will be given in a word document format easily readable on any computer.

The PHASE project will use standard vocabularies for all data types present to allow inter-disciplinary interoperability. If it is unavoidable to use uncommon or generate project-specific vocabularies, i.e., WP5 and the qualitative data, clear descriptions will be provided to facilitate the easy re-use of the data. The description of data and metadata vocabularies, standards, and methodologies refer to points 2.1 and 2.2.

2.4. Increase data re-use (through clarifying licenses)

In order to create clarity for potential users towards the use of the data, suitable licenses will be assigned to the data, using creative commons licenses (mostly CC-BY).

No embargo will be applied beyond those that are potentially requested by social science journals. All the data produced and used in the PHASE project are useable by third parties, particularly after the end of the project.

Once delivered to the European Commission and approved, the data files will also be made public on the Fellow’s website by linking published papers and project data to the SocArXiv repository. To increase the potential of re-use, the data will be actively promulgated on social media.

The exact time when the data will be deposited on the platform depends on the amount of time required to prepare the data and the documentation. The data will remain reusable without any specified time limit under the Creative Commons license.

Data quality assurance is provided in line with the principles of research integrity and responsible research. Data management will be overseen by the host organization — Bocconi University — with considerable experience with EU-funded FP7 and H2020 projects.

3. Allocation of resources

Costs related to open access to research data are eligible costs as part of the Horizon 2020 MSCA grant specified in the Grant Agreement. The project will provide easy access to the research data using SocArXiv, a free online repository. Depositing materials on SocArXiv will not generate any costs, and all materials will be prepared and checked by the Fellow, potentially in collaboration with other researchers who will join the project on a case-by-case basis as co-authors. Data management will be the responsibility of the Marie Curie Fellow and project coordinator.

All publications resulting from the project will be published as open access articles in peer-reviewed journals. The costs related to open access publications (article publishing charges – APC) are eligible for Horizon 2020 MSCA grants under the budget category Research, training, and networking costs. There will be no further costs related to the preservation of the data beyond the funding period.

All output generated in the project will be open for re-use and sharing without any restrictions. The project's main dissemination and communication objective is to popularize its findings and outcomes among the general public and policymakers.

4. Data security

Raw data will be stored on secure university fileservers, and backup versions will be saved on external hard drives and personal computers of the Fellow and other researchers involved in the project. For the project's duration, the research data master files will be stored on the university fileserver with the Fellow's partner institution to ensure long-term and secure storage. Backups will be made from the master file location and stored on local drives – on personal laptops with responsible researchers. Working copies will be accessible on cloud storage (Dropbox), enabling researchers to access the data and allow editing environment. The updated working copies will be synchronized regularly (after every edit) with the master copy location.

Dropbox, Bocconi University, and SoxArXiv use state-of-the-art technologies for secure data storage and transfer and manage the rights of the users. The data collected for the project will be fully anonymized, and no potential data user will be able to identify interviewees based on the information recorded in the data. The data will be stored in a minimum of two separate locations to avoid data loss. The data will not be transferred on USB drives. All data will be labeled systemically in order to ensure coherence and transparency.

5. Ethical aspects

The proposal does not pose any ethics issues. There is no need to apply for further ethics approval or other ethics documents. The Fellow and the supervisor are aware of and will comply with ethical principles and applicable International, European, and national law, based on the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights and its Supplementary Protocols, and the UN Convention on the Rights of Persons with Disabilities (UN CRPD), and other relevant ethics rules and regulations.

The data used during the project are publicly available and can be freely used for the project. Previously collected personal data by the EU (ESS) is anonymized and processed according to EU regulations. The research is performed inside the EU; no data will be transferred from the EU to non-EU countries or international organizations. The staff involved in the project will work in Milan, at Bocconi University, no health and safety issues are involved. The research project has an exclusive focus on civil applications.

Specifically, the project does not involve: research on human embryos and fetuses, research on human stem cells or tissues, collecting new personal data, or research involving human participants (collection of biological samples, personal data, medical interventions, interviews, observations, tracking, or the secondary use of information provided for other purposes), processing of personal data (ESS data are anonymized), research on animals or animal experiments, research on non-EU countries, the use of elements that may cause harm to the environment, to animals or plants, dealing with endangered fauna or flora, protected areas, dual-use items in the sense of Regulation 428/2009, or other items for which an authorization is required, or potential for misuse of research results.

6. Other issues

The project does not make use of other national / funder / sectorial / departmental procedures for data management.

Annex

Project summary

The recent surge in populism marks a new phase of European politics, posing fundamental challenges to scholars and the policy community. Some argue that existing nationalist and authoritarian attitudes are driving populism. Others posit that it is rooted in liberal institutions' failure to deliver equal life chances. Once in power, populists tend to discriminate against precarious communities and question the role of expertise in governance. However, there is no systematic cross-country research on the bi-directional association between ill health and populism. The project contributes to the existing knowledge in two ways. Firstly, it crosscuts scholarly polarizations on the causes of populism by posing the original question of whether individuals' ill health might be an early warning sign for the deteriorating health of democracies in Europe. Secondly, it extends the literature on the impact of populist governance by asking who the winners and losers of populists in power in terms of health in Europe are. In answer to the first question, the project creates a longitudinal, hierarchical cross-country panel dataset nesting individuals in EU regions and measures the impact of ill health on populist support after 2000. In answer to the second question, the project adopts a coding scheme for populist governments in Europe and assesses the variegated impact of populism on peoples' health using dynamic multilevel modeling. Finally, through a small-N case study, the project analyses the populism – ill health mechanism in detail. The project offers theoretically innovative and methodologically rigorous, cutting-edge contributions to sociology, political science, and public health and the public debate on populism through a set of outreach activities.

History of Changes

Version	Date	Change
1.0	June 24, 2021	▪ Initial version