**MANNHEIM MASTER OF APPLIED DATA SCIENCE & MEASUREMENT**

**MASTER PROJECT PROPOSAL**

|  |
| --- |
| **Name of student(s)** |
| Deliverance Bougie  Stefan Zimmermann |
| **Name of supervisor and Email address. Please note: Your supervisor must be either a Professor or Honorary Professor, in accordance with §9 Paragraph 1 of the Exam Regulations.** |
| Prof Dr Florian Keusch, [f.keusch@uni-mannheim.de](javascript:linkTo_UnCryptMailto('nbjmup%2Bg\%2FlfvtdiAvoj.nbooifjn\%2Fef');)Prof Dr Alexandru Cernat, [alexandru.cernat@manchester.ac.uk](mailto:alexandru.cernat@manchester.ac.uk) |
| **Abstract** |
| With its very [broad range of topics](http://companion.soep.de/Topics of SOEPcore/index.html), the [many different questionnaires](http://companion.soep.de/Survey Design/SOEP Questionnaires.html), the household context, the long duration of the panel and the many [special sample groups (migrants, refugees, high income households, East Germans)](http://companion.soep.de/Target Population and Samples/index.html).comparisons over time can be presented with the SOEP very well.  In order to make this inventory of social science data and research results as widely accessible as possible and to establish a new culture of using scientific results in other areas of society as well, our project aims to create a platform that allows non-scientists such as journalists or historians to work indirectly with SOEP data without needing direct data access. So the master project would establish a transfer interface (with shiny) with aggregated data so that non-scientists can use this dashboard to answer current societal challenges outside the social science discipline. |
| **What is the (working) title of the Master project proposed?** |
| Development of a data visualization platform for non-scientists with SOEP data. |
| **What is the key question?** |
| How to create an interactive interface that provides information (basic statistical indicators) to the non-scientific community in a user-friendly way, so that simple analyses are possible. |
| **Which data source(s) do you use to answer this question?** |
| The [Socio-Economic Panel (SOEP)](https://www.diw.de/sixcms/detail.php?id=diw_01.c.814095.en) is one of the largest and longest-running multidisciplinary panel studies in the world and has been providing cross-sectional and longitudinal data [exclusively to researchers](https://www.diw.de/en/diw_01.c.601584.en/data_access.html). So, although in the age of Fake News there are data sources like SOEP that can answer many social, economic, and societal questions and enlighten society, access to SOEP microdata is only available to researchers and for teaching purposes. Helpful information for an enlightened society is usually only accessible to a small part of society. |
| **Is the data you use already available? If so, where? If not, which data collection methods do you apply?** |
| We would need to apply for data access with SOEP. Since Deliverance and I are part of a master's thesis/master project at an university, we need a project-based data sharing agreement between the university and SOEP:  <https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.88926.de/soep_application_contract.583953.pdf>  This is how this process works:  [https://www.diw.de/en/diw\_01.c.601584.en/data\_access.html#c\_741347](https://www.diw.de/en/diw_01.c.601584.en/data_access.html" \l "c_741347)  **The responsible project head have to fill out the required document and request the data. We would like to have the 100 percent SOEP EU v36 edition.** |
| **Which methods of data preparation, analysis, and evaluation do you plan to apply?** |
|  |
| **Which skills do your team members need to work on this project?** |
| * Knowledge of statistical localization and dispersion parameters * Programming skills in data preparation in R and Stata * Understanding about metadata and dataset documentation * Programming skills in shiny to build interactive dataplatform. * Knowledge and programming skills in the area of data visualization |
| **What are the general tasks of each team mate?** |
| Since the project is very ambitious and requires a lot of programming, all project members will have to work on these areas. However, splitting the work packages as soon as the idea is clearly defined is possible at any time.   * Data preparation in Stata * Generation of aggregated tables in csv format * Processing of aggregated tables in plotly graphs * Creating the user interface in shiny |