Plan Overview

A Data Management Plan created using DMPonline

Title: Beyond Boardrooms: The Impact of Corporate Quota on Gender Equality in Organizations

Creator: Zoltan Lippenyi

Principal Investigator: Zoltán Lippényi

Data Manager: Zoltán Lippényi

Project Administrator: Zoltán Lippényi

Affiliation: Rijksuniversiteit Groningen

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0001-8751-5741

Project abstract:

Women in large European companies occupy less than 30% of corporate leadership positions and the pace of advancement of gender diversity in leadership is decelerating. European governments have experimented with corporate boardroom quotas to ensure a more equal representation of women in leadership positions, hoping that this representation would trickle down to affect women's career opportunities at lower organizational ranks and spill over to organizations not directly targeted by these policies. It is unknown to what extent and in what ways quotas are successful in achieving broader equality aims. Beyond Boardrooms will be the first research project to address in what ways quota schemes impact gender equality. I will investigate how quotas impact women's access to corporate leadership positions and HR policies and practices, to what degree the impacts trickle down to earnings inequality between men and women and occupational sex segregation, and to what degree they spill over across organizations, by examining a previously unresearched context, the Netherlands. The study will consider the impact of two quota regulations, the corporate diversity target (2013-2020) and corporate quota (2022-2024). I will construct the Netherlands Integrated Database of Inequality in Organizations (NIDIO), a unique infrastructure that combines (1) administrative register and survey data about the earnings, occupation, and employment histories of Dutch organizational leaders and employees, and (2) business registers and diversity monitor surveys about organizational HR practices and policies. This data will be used by the Beyond Boardrooms project to provide important insight into the impacts of past and ongoing regulatory efforts as well as guidance for improving quota regulations and corporate strategies to further gender equality.

ID: 111771

Start date: 01-01-2023

End date: 31-12-2027

Last modified: 12-12-2022

Grant number / URL: VI.Vidi.211.231

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Beyond Boardrooms: The Impact of Corporate Quota on Gender Equality in Organizations

General Information

Name applicant and project number

Name applicant: Zoltán Lippényi
Project number: VI.Vidi.211.231

Name of data management support staff consulted during the preparation of this plan and date of consultation.

Marieke Steenman (m.steeman@rug.nl)

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

The project will re-use digital register data containing information on Dutch companies and their employees and already collected survey data from the Dutch Central Bureau of Statistics's System of Social Statistics Datasets (**SSD**). From SSD, the project will use
1) yearly registers of the Dutch population (e.g., vital registers, education), until 2022
2) registers of the population of Dutch companies and their functionaries, until 2022
3) Dutch Labor Force Surveys (**LFS**; 1999-2022), a representative rotating panel survey of Dutch households collected by the Central Bureau of Statistics

In addition, the project will use the Company Monitor of Women in Leadership (**CMWL**;2003-2022), a representative panel survey of large Dutch companies commissioned by the Ministry of Education, Culture, and Science of the Netherlands and collected by Bureau Pouwels.

Access to SSD data is granted to authorized research projects via Remote Access by the Central Bureau of Statistics. CMWL is publicly available via DANS. The project will request to link CMWL with SSD (see points 3 and 4 for further information).

All raw and processed data are stored and analyzed on the CBS servers (as SPSS/Stata data files), and CBS provides internal storage for projects to store processed data (up to 100GB). The estimated volume of processed data and results stored will be about 40GB.

SSD data are pre-processed by CBS (quality control and anonymization), and generated outputs are subject to output control to avoid identifiability.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

The project will gather publicly available information published on the SER Diversity Portal (SER Diversiteitsportaal,**SERD**, starting 2023 January) from company diversity and inclusion reports. Digital tabular data will be created at the company level (c. 5000 companies, volume: 50MB) by coding the company reports. Specifically, the following information will be recorded:
1) company gender diversity targets (leadership and below the top)
2) existing and planned diversity and inclusion policies and programs
3) company self-evaluations of reaching targets (years 2023/2024)

The project members (PI and 2 PhDs) will perform the data collection and coding using double-coding to quality and accuracy of the codes. The resulting dataset (.csv and .sav) will be linked to CBS SSB using unique Chamber of Commerce company identifiers (see further points 3 and 4), and an anonymized version will be deposited at DANS with accompanying documentation of coding schemes upon the completion of the project (2027) and after a potential publishing embargo. Preparations for and the deposition of **SERD** will be performed via DataverseNL and in consultation with data experts from the University of Groningen Research Data Management office.

1.3. How much data storage will your project require in total?

- 10 – 100 GB

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

Using the datasets mentioned under 1., the project will create the Netherlands Integrated Database of Inequality in Organizations (**NIDIO).** NIDIO is an analyzable linked employer-employee dataset consisting of survey and register data for the purpose of studying wage and employment inequalities within Dutch organizations. A working paper and documentation of NIDIO will be published and deposited to DANS. The deposition will (broadly) consist of
1) textual descriptions of the steps taken to link register datasets and the justification of steps (e.g. loss of coverage)
2) coding schemes to map administrative concepts in registers to research analytical concepts and their justification
3) preparation of data for analysis (e.g. weighting)
4) open-source statistical software syntax (R)
5) codebook containing names and paths of component datasets on the CBS server, variable names, their content, and how they were processed.
The deposition will not include raw or processed microdata as these can only be accessed by CBS-authorized research projects. Therefore, the aim of NIDIO and the accompanying meta-data is that researchers can re-use and adapt the code to their own applications with the same of similar linked data. Furthermore, researchers authorized by CBS to use the component datasets can compile them and replicate results.
The deposition will be performed via DataverseNL.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

The NIDIO meta-data will follow DANS metadata guidelines and steps will be taken to improve findability and searchability (e.g. using CMDI). Preparations for the deposition will be performed in consultation with data experts from the University of Groningen Research Data Management office.

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

A designated folder on the Groningen Networked Research Storage (Y: drive) will be used to store and share project information among project members. Project-generated outputs (e.g. meta-data and collected data) will also be stored here. Data files provided by CBS (e.g., SSD) are stored on the CBS server.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Additional security measures (please specify)

Default security measures of the University of Groningen apply to the institution's networked research storage. Data stored on the University of Groningen storage does not include personal data. Provisions regarding specific datasets:
Linking of the **CMWL** dataset with the CBS SSD will be performed by CBS. As agreed with Bureau Pouwels (collector of the dataset), they provide the original dataset upon consent from the Ministry. CBS links and anonymizes the dataset before making it accessible for the VIDI project via remote access. Therefore, the VIDI project members will not get access to and store the original survey data that includes company identifiers.
The **SERD** dataset includes self-disclosed and publicly available company data (on the SER portal). This data will be deposed after

anonymization according to European Data Protection Board (EDPB) and Universiteit van Nederland GDPR guidelines.
Individual and company data (**SSD**) stored on the CBS server are subject to additional CBS security protocols specified by the law. The CBS implements these measures, and the project will follow the CBS guidelines regarding data protection.

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data be ensured?

- No

4.2 How will ownership of the data and intellectual property rights to the data be managed?

Existing register and survey datasets within the SSD (see 1.) are owned by CBS. The CMWL survey (publically available via DANS) is owned by the Ministry of Education, Culture, and Science of the Netherlands.
The University of Groningen will own and retain intellectual property rights over the newly collected SERD. SERD will not contain confidential data and it will be made publicly accessible via DANS.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- No

Upon project completion, the NIDIO meta-data and the SERD dataset (after anonymization) will be deposited to DANS without restriction. If there are paper submissions and unfinished doctoral dissertations, a short embargo period (max. 2 years) will be installed.

5.3 What data will be made available for re-use?

- Other (please specify)

All data will be made available, except data from the SSD (see point 1) owned by CBS and subject to research authorization requirements.

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available after completion of project (with embargo)

Replication packages to articles (code and exact reference to CBS data to allow replication for authorized researchers) will be released upon publication of papers. NIDIO meta-data and SERD will be available after the completion of the project (and reasonable embargo, see 5.2)

5.5 In which repository will the data be archived and made available for re-use, and under which license?

The deposition will be made to DANS via DataverseNL.

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

Open-source software (R) will be used for the project.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Costs of accessing and using the CBS data (e.g., authorization, data linking) are specified in the project budget (in total: 40,800 euro). No additional charges are expected for the deposition of (meta)-data. In case technical expertise or advice is needed for the deposition, the project has access to consult (free of charge) from data experts from the University of Groningen Research Data Management office.

The project members (PI + 2 PhD researchers) dedicate 1,5 years of researcher time (10 % the total 11,5 years of research time) to the preparation of the meta-data NIDIO, the SERD data collection, documentation, and data depositions (in addition to data management work for research)