Plan Overview

A Data Management Plan created using DMPonline

Title: Modeling Automation Maturity in Software Development

Creator: Simcha Vos

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2021)

Project abstract:

Analyzing automations in order to recommend to developers what they should automate next.

ID: 166599

Start date: 01-10-2024

End date: 31-05-2025

Last modified: 15-12-2024

Modeling Automation Maturity in Software Development

0. Administrative questions

I. Data description and collection or re-use of existing data

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Automation quantitative data	csv	Aggregated from public github repositories (so no raw data, just accumulated)	Obtaining quantitative data to be able to build taxonomy	Project storage	Me and responsible researcher
Survey results	-	Survey	Obtaining research results (taxonomy of automations)	Qualtrics and Project storage (after survey finished and removed from Qualtrics)	Survey owner / Me and responsible researcher
Feedback results	-	Responses on GitHub issues	Validating research output in practice	Project storage	Me and responsible researcher

- < 250 GB

II. Documentation and data quality

- Data will be deposited in a data repository at the end of the project (see section V) and data discoverability and re-usability will be ensured by adhering to the repository’s metadata standards
- Data dictionary explaining the variables used
- Methodology of data collection

III. Storage and backup during research process

- Project Storage at TU Delft

IV. Legal and ethical requirements, codes of conduct

- Yes

HREC is submitted

- No

- No, I will not work with any confidential or classified data/code

Data will be openly accessible.
In case licenses are applicable, they will be honored, in case of analyis on GitHub repositories without a license, it is copyrighted. All repositories that are analyzed, are public GitHub repositories. We do not alter or distribute any files. We simply analyze and aggregate some statistics on the source files.

V. Data sharing and long-term preservation

- All validated non-positive results
- All data (and code) underlying published articles / reports / theses
- All data (and code) produced in the project

- I will share my data and code via git(lab)/subversion and also create a snapshot in a repository
- All data will be uploaded to 4TU.ResearchData

- < 100 GB

- At the end of the research project

- BSD
- CC BY

VI. Data management responsibilities and resources

- Yes, the only institution involved

Assistant Prof Sebastian Proksch s.proksch@tudelft.nl at SERG

4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. We do not expect to exceed this and therefore there are no additional costs of long term preservation.