# Read me

This repository contains supporting materials used in the research conducted for the PhD thesis ‘*Public information infrastructures’* at the TU delft, written by Wendy van Donge.

The repository includes interview and survey question sets as well as Python scripts used for hierarchical clustering analyses. No original research datasets are included due to confidentiality and privacy considerations.

## 1. Interview Question Sets

Three files containing interview questions used during phases 2 and 3 of the research:

| **File Name** | **Language** | **Format** |
| --- | --- | --- |
| *Interview questionair\_Phase 2\_Quality\_ NL.docx* | Dutch | Word |
| *Interview questionair\_Phase 2\_Design choices\_NL.docx* | Dutch | Word |
| *Interview questionair\_Phase 3\_Configurations\_NL.docx* | Dutch | Word |

## 2. Survey Questionnaires (Qualtrics)

These surveys were developed and administered with Qualtrics. They are provided in PDF format.

| **File Name** | **Language** | **Format** |
| --- | --- | --- |
| *Vragenlijst Online Survey\_phase 3\_ Configuraties\_NL.pdf* | Dutch | PDF |
| *Vragenlijst Online Survey\_phase 3\_ Quality\_NL.pdf* | Dutch | PDF |

## 3. Python Scripts for Hierarchical Clustering

This repository also includes three Python scripts designed to perform hierarchical clustering on different sets of variables:

| **Script Name** | **Variable Category** | **Format** |
| --- | --- | --- |
| *Phython\_Clustering\_Governance.py* | Governance-related variables | Phyton |
| *Phyton\_clustering\_DataFormat.py* | Data structure / formatting variables | Phyton |
| *Phyton\_Clustering\_Technal.py* | Technical variables | Phyton |

**Instructions for Running the Python Code**

* **Important:** The underlying dataset is not provided.
* To run the scripts, users must supply their own dataset.
* In each Python script, lines that require user modification are marked with **[to do]**.
* Relevant variable names and dataset paths must be adapted to match the user’s data input.
* **Please note:** Attention is required when specifying variable types (e.g., binary, categorical, continuous) to ensure that the clustering method treats the variables appropriately.

## Privacy Note

The original research data are not shared due to privacy, confidentiality, and ethical research obligations. The repository only contains non-sensitive materials suitable for open distribution.