**Explanatory Memo regarding data files appended to the submission *A replication study for good practices in Quantitative Text Analysis in Public Administration***

This data folder contains 4 subfolders:

1. **Data**: this folder contains all data that is needed to replicate the study. These are 2 .csv files (can be opened with among others Microsoft Excel). *coded-tweet-data.csv* contains the labeled tweets that serve as training set and test set to the machine learning model. *agency\_tweets\_database.csv* contains unlabeled tweets. **N.B.** Both files are identical to the files appended to Anastasopoulos and Whitford (2019).
2. **rmd\_files:** This folder contains 3 R-markdown files[[1]](#footnote-1). These files can be opened and run with R-studio[[2]](#footnote-2). Executing these files will replicate the entire study. If the data-files are kept in their folder after extraction of the zip, the script is able to find them without any alterations. Note that for the script to run properly, some extra R-libraries need to be installed, preferably the versions indicated in the R-markdown files.
3. **html\_files:** This folder contains the content of the rmd-files (including output) in HTML-format. These files can be opened with any web browser (firefox or google chrome are recommended.). These files contain no new information regarding the rmd-files, but are meant for people that are not interested in running the code themselves.
4. **Figures:** this folder contains all the figures created with the scripts. When the rmd-files are run after extraction of the zip-file, all figures will be placed in this folder.

1. See for more information about R-markdown: <https://rmarkdown.rstudio.com/lesson-1.html> [↑](#footnote-ref-1)
2. <https://rstudio.com/> [↑](#footnote-ref-2)