**Pfizer readme p\_meta D3**

The specifications outlined in SAP are translated into a set of CSV tables, which are stored in the p\_meta\_data folder located at https://github.com/UMC-Utrecht-RWE/Pfizer/tree/Interim-3-development/Data%20characterisation/PfizerScript/p\_meta\_data. These tables are essential for the script, as they contain information that is imported and used by the script.

If there are changes in the specifications, most of these changes can be made by updating the relevant files in the p\_meta\_data folder. Additionally, if data is missing in the script's output for a specific DAP, the first thing to check is if the metafiles are filled correctly.

The following files are essential and needed for the translation of the SAP (in the green block in the figure at page 2):

* Pfizer\_study\_variables.csv
* Pfizer\_algorithms.csv
* Pfizer\_dictionary.csv
* Pfizer\_scores.csv
* (Pfizer AESI\_information)

The following files are essential for the extraction of the concepts/study variables (in red):

* Pfizer\_additional\_concepts.csv
* 20221208\_ALL\_full\_codelist.csv
* 20221208\_ALL\_drug\_proxies\_codelist.csv
* Pfizer\_vaccines\_codelist.csv

The Program.csv file provides an overview of the steps in the script (in orange).

During the script's execution, the metafiles are analyzed, and the necessary information is extracted and stored in global variables or in .rds files in the folder Pfizer\Data\characterisation\PfizerScript\g\_intermediate\tmp. This is done in the following steps:

* Step\_00\_SetParameters.rds
* Step\_00\_SetCodeSheets.rds

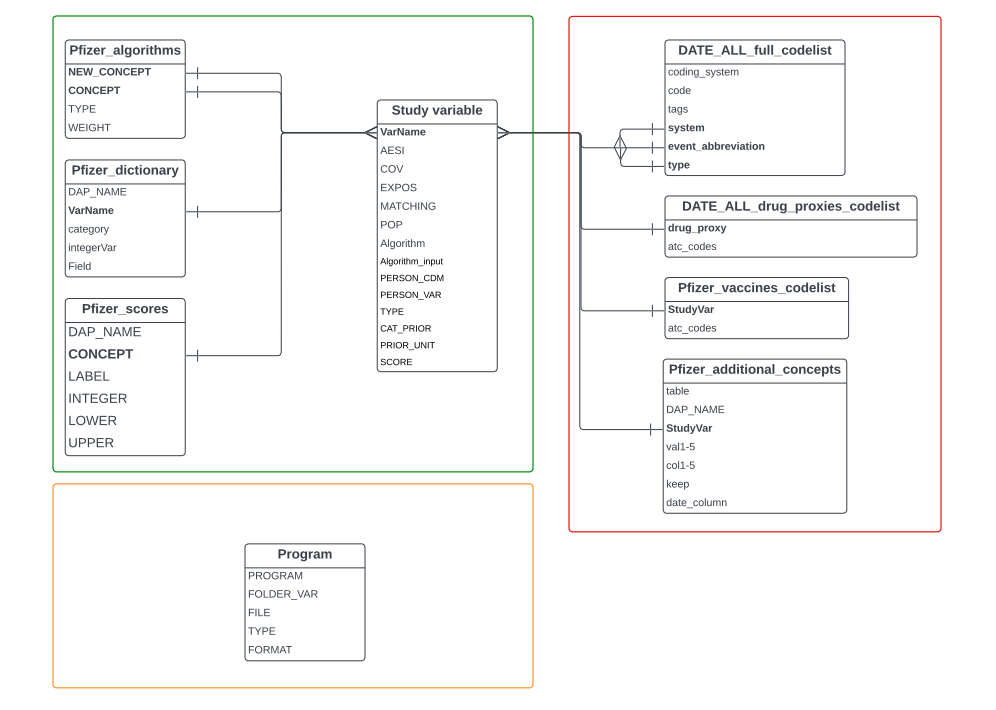
Note that when running the script in multiple stages, the two aforementioned scripts must always be rerun first, as they create the necessary variables and R objects that are used throughout the script.

There is a need to create a script that checks the assumptions of the metafiles. While some checks are already performed in Step\_00\_SetParameters.rds, additional checks can be added to ensure the integrity of the metafiles.

Please refer to appendix 1 for a diagram that subdivides the script into multiple stages. However, note that the visualisation may not be 100% accurate due to the complexity of the process.

**Model**

The figure below illustrates all the tables mentioned on page 1, along with their relevant variables. The lines indicate how the tables are connected. It's important to note that the files and related script were developed piece by piece over a long period of time, so the naming and relations may not be perfect. With current knowledge, a more structured model would be developed accordingly. Nevertheless, the model depicted here contains 80% of the study information, and the D3 script is built upon this layer.



**Advise**

It is recommended to avoid using Excel when working with CSV tables from the model. Instead, it is more suitable to use Notepad or Notepad++ as saving the file in Excel can unintentionally alter the file's formatting and introduce unwanted spaces, leading zeros, or strange characters. While Excel may be more user-friendly, it is advisable to open the file in Notepad after making all the changes in Excel and review it for any spaces, strange characters, or leading zeros.