**Summary**

This document will explain the idea of sorting and saving the data from Bühlot. It gives an overview so everyone, even later on, can work on this project.

First of all the data must be safed with the correct names for each station. So the process starts by collecting the data out in Bühlertal. Here is a list of all the names and how they must be named:

* Niederschlagsmessungen:
  + Butschenberg
  + Grundigklinik
  + Hundseck
  + Schafhof
  + Schönbrunn
  + Schwabenquelle
  + Sportplatz
  + Sternenberg-Schlammfang
  + Winterberg
* Pegel:
  + Pegel1\_Büchelbach
  + Pegel2\_Büchelbach
  + Pegel4\_Büchelbach
  + Pegel1\_Bühlot
  + Pegel2\_Bühlot
  + Pegel4\_Bühlot
  + Pegel1\_Schwabenbrünnele
  + Pegel2\_Schwabenbrünnele
  + Pegel4\_Schwabenbrünnele
* Bodenfeuchtemesser:
  + Schafhof1\_Table1
  + Schafhof1\_Table2
  + Schafhof5\_Table1
  + Schafhof5\_Table2
* Tensiometer:
  + Schafhof\_Tensiometer
  + Sprengquellen\_Tensiometer\_oben\_nord
  + Sprengquellen\_Tensiometer\_oben\_sued
  + Sprengquellen\_Tensiometer\_unten\_nord
  + Sprengquellen\_Tensiometer\_unten\_sued

It is important that they will be saved like this because the python code “data\_preprocessing” - which also will be found in this folder - will preprocess the data by the names. If a file is named differently the data will not be preprocessed. An overview of all the different stations is summarized in an excel table called “metadata” – which also will be found in this folder.

The raw data will be found in the folder “data”. There it is sorted by each year and then by the date of collection. The file names in this path are slightly different than the names you will find in the network folder (Network location: HD-FILES KIT (R:) > hiwi > Bühlot > Rohdaten > Daten Bühlot). The content is the same, but the names are different. The reason for that is because in the past few years there were different people collecting the data and each person gave the files different names. To create a uniformly construction the names were changed. If later a problem with names will be found or any other one it might make sense to check the files in the network path because this is the folder with the original files.

The purpose of the python code “data\_preprocessing” is to read in all the collected data, sort it by their different variables and then safe it in the folder named “data\_Export”. What the code does and how it works is explained in the code.

Tensiometer have three variables: ground water level, logger temperature and water temperature. If you go to the folder “data\_export” (data > data\_export) you will find all the preprocessed variables. In each folder of the three variables of a tensiometer you will find files that are called “alt”. I labeled them old because a few years ago there were different files of each tensiometer station. I don’t really know what exactly they are. I have been ignoring them so far and only used the other ones, because for those I can surely say that these are correct files.

Same goes for the variables “river water level 1”, “river water level 2”, “river water level 4”. I have been ignoring those too because I am not sure what they really are. I know that a Pegel station collects these three variables: water level, water temperature and water conductivity. But I don’t know which file is which variable, so I just called them ground water level 1 to 4.