Manual

Guidance on data storing for monitoring and assessing the cropping systems

Work Package 5 of the EU-project SoilCare

Ioanna Panageaa, Jan Dielsa , Jean Poesena, Guido Wyseurea

aKU Leuven, Leuven, Belgium

# Foreword

The aim of this report is to provide details and guidance to the Study Sites partners about the way the experimental data should be stored. The data collected through the excel files provided will be stored in a safe guarded cloud storage database in formats that facilitate comparison and can be consulted by the project members. Data sharing and use of shared data will be done in accordance with what has been agreed in the SoilCare Consortium Agreement and in the Agreement on data sharing. An overview of the WP5 related activities, milestones and deadlines is also included.

We would like to thank the study site teams and work package leaders for their comments and feedback.

*Ioanna Panagea, Jan Diels, Jean Poesen Guido Wyseure,*

October 2018

# Overview of the WP5 related tasks

|  |  |  |  |
| --- | --- | --- | --- |
| Month | Date | To do, or to provide to WP5 | To receive from WP5 |
| 31 | sep-18 | Follow-up with the Study-Sites | WP5- Follow up of the monitoring1.Excel sheet for the CS monitoring 2.Excel sheet for the historical existing weather data and future weather data3.Document on the way to fill the excel sheets for monitoring |
| 32 | oct-18 | Filled information on the excel sheet about the experiments’ set up, observations of the 2017-2018 cropping season if available. and historical weather data (till end August 2018) for testing the database and keep track of the first cropping season if applicable.  |  |
| 33 | Nov-18 |  |  |
| 34 | Dec-18 |  | WP5: quality checked database to WPL and SS |
| 35 | Jan-19 |  |  |
| 36 | Feb-19 |  |  |
| 37 | Mar-19 |  | Improved excel to the SS (with the modifications after the quality check)  |
| 38 | Apr-19 | SS are doing the soil sampling campaign for 2018-2019 cropping season as agreed in Billund meeting |  |
| 39 | May-19 |  |  |
| 40 | Jun-19 |  |  |
| 41 | Jul-19 | Filled information on the excel sheet about the experiment measurements for updating the database with the results from the last cropping season. |   |
| 42 | Aug-19 |  |  |
| 43 | Sep-19 |  | Follow up of the monitoring |
| 44 | Oct-19 | Filled information on the excel sheet about the weather data (September 2018 till end August 2019 |  |
| 45 | Nov-19 |  |  |
| 46 | Dec-19 |  |  |
| 47 | Jan-20 |  | WP5: quality checked database to WPL  |
| 48 | Feb-20 |  | WP5: demonstration events? |
| 49 | Mar-20 |  |  |
| 50 | Apr-20 | SS are doing the soil sampling campaign for 2019-2020 cropping season as agreed in Billund meeting | D5.2 Deliverable demonstration activities in Study Sites |
| 51 | May-20 |  |  |
| 52 | Jun-20 | Filled information on the excel sheet about the experiment measurements |  |
| 53 | Jul-20 |  | D5.1 Deliverable database with monitoring data |
| 54 | Aug-20 |  |  |
| 55 | Sep-20 | Filled information on the excel sheet about the weather data (September 2019 till end August 2020 |  |
| 56 | Oct-20 |  |  |
| 57 | Nov-20 |   | D5.3 Deliverable Monitoring results  |
| 58 | Dec-20 |  |   |
| 59 | Jan-21 |   |   |
| 60 | Feb-21 |   |   |

# Goal

WP5 aims to make the experimental data readily available to the partners in a way that information is useful, easy to access and download and at the same time well protected. In the beginning, the database modelling includes UML diagrams to provide a standard way to visualize the database design. After the design was finalized, these excel sheets were created for the study site partners to collect the data which will be uploaded in the SoilCare database. The database will ensure the consistency, easy retrieval, update, manipulation and maintenance of the data. Then, as the third component a web-interface or web application will be designed which will enable the users to access, visualize, download and query the database.

The goal of these excel files is to collect all the information required to be able to monitor, process and analyse the Cropping Systems (CS) of the SoilCare Study Sites (SS) in an standardised way that is compatible with the database structure. All the information required for monitoring a Cropping system properly, and not only experimental results, are going to be collected by WP5 and stored in a safe guarded cloud storage database, and can be consulted by the project members through the online application of the database.

Two different files will be used for the monitoring of the SS cropping systems. The first one entitled **SoilCare\_experiments\_monitoring\_form\_vX** will include all the information required for the experiments, treatments and observations as described in the Report 10, D4.2 Monitoring Plan for the Study Sites. Details can be found in **Part 1** of this report. The second one entitled **SoilCare\_weather\_monitoring\_form\_vX** will be used for monitoring the climatic and weather data separately. Details can be found in **Part 2** of this report.

# Part 1: Document to monitor the experiments and cropping systems

## Terminology used

**Experiment:** An experiment is a procedure carried out to support, contradict, or validate a hypothesis. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated. Experiments vary greatly in goal and scale, but always rely on a repeatable procedure and logical analysis of the results. In physical sciences, typically they focus on the replication of the different treatments to determine whether they result in statistically significant differences or not.

The validity of an experiment is directly affected by its construction and execution, thus attention to experimental design is extremely important.

**Experimental design:** Experimental design is the process of planning a study to meet specified objectives. Planning an experiment properly is very important in order to ensure that the right type of data and a sufficient sample size are available to answer the research questions of interest as clearly and efficiently as possible. Most experimental designs require experimental units to be allocated to treatments either randomly or randomly with constraints, as in blocked designs.

! Details for common experimental designs can be found in the “Experimental Design Details” section of this guide.

**Factors**: A factor of an experiment is a controlled independent variable, whose levels are set by the researcher. Level is a specific state or value (qualitative or quantitative) of a factor. The levels of greatest interest should be clearly defined for each primary factor (the factors whose effects need to be quantified). For example, if the experimental units are tilled up to 10 cm, 20 cm, 40 cm, those depths would be three levels of the factor tillage. Typical factors are type of tillage, amounts of fertilizer, addition (amount and/or type) of organic matter like compost, farm year manure or other, type of cover crops, etc. Each factor has two or more levels.

**Treatment:** is an unique combination of factors level. Or in other words a specific system that researchers administer to experimental units i.e. plots and whose effect is to be measured or observed. An experiment consists of several treatments. In a factorial treatment, the treatment structure consists of a unique combination of factors that the researcher wants to study and about which the researcher will make inferences. So every treatment represents an unique combination of different levels of different factors. The primary factors are controlled by the researcher and are expected to show the effects of greatest interest on the response variable(s). The levels of greatest interest should be clearly defined for each primary factor. The levels of the primary factors represent the range of the inference space relative to this study. The levels of the primary factors can represent the entire range of possibilities or a random sub-set.

**Replication i.e. repetition:** If a specific treatment appears more than one time, it is defined to be replicated. True replication refers to responses that are treated in the same way. They should yield similar results. In other words, researchers would want to ensure consistency in their results. The simplest way to achieve consistency in the results is through the replication of the same treatment on several plots or experimental units. Repetition on the same plot is not recommended, as observations are not independent. Moreover, the use of several small plots instead of one large plot ensures minimization of the effect of uncontrolled variability in the field.

**Plot i.e. experimental unit:** is the smallest unit receiving a certain treatment. The information or-data for comparison are from such single units. A plot has a treatment level and belongs to a block (blocked design) or to a main plot (split plot)**.** Every plot belongs either to a block or to a main plot. The block or main plot belongs to field (i.e. location). All blocks/ main plots could belong to the same field or could be on different fields.

**Blocks:** Are groups of experimental units that are formed to be as homogeneous as possible with respect to the block characteristics. The term block comes from the agricultural heritage of experimental design where a large block of land was selected for the various treatments, which had uniform soil, drainage, sunlight, and other important physical characteristics. Homogeneous clusters improve the comparison of treatments by randomly allocating levels of the treatments within each block.

**Field i.e. location:** is the physical location where the experiments are performed. In this database, a field is characterized by the previous history, landscape (topography, shape etc.), inherent soil properties (soil profile characterization etc.), and climate (historical weather).

**Control treatment:** is a “standard” treatment that is used as a baseline or basis of comparison for the other treatments. This control treatment might be the treatment in common use, or it might be a null treatment (no treatment at all). For example, a study on the efficiency of fertilizer could give some experimental plots no fertilizer at all or the common fertilizer used.

The control is also considered as a treatment

## Experimental Designs Details

The scientific experiments are generally classified into two categories. Single-factor and multifactor experiments. In a single-factor experiment, only one factor varies while other conditions are kept constant. In these experiments, the treatments consist of different levels of the single variable factor. In multi-factor experiments (also referred as factorial experiments), two or more factors vary simultaneously. The treatments in factorial experiments consist of two or more levels of the two or more factors. The multi factorial treatments can be applied to any of the designs with various arrangements.

### Randomized designs

Three major experimental designs commonly used for both types of experiments (multi or one – factor) are:

- completely randomised (CRD)

- randomised complete block (RCB)

- latin square (LS)

An example of the different arrangements of the plots of a 2x2 factorial experiment is presented (Figure 1) to help the SoilCare Study Site partners to categorize their experiments in a uniform way.

The treatments of the example experiment are replicated 4 times (4n). The experimental plots can be arranged in different combinations leading to various experimental designs-arrangements (Figure 1).



Figure 1: a) completely randomized design (CDR), b) a randomized complete block design (RCB), c) a Latin square design (LS), d) a split plot- whole plot completely randomized (Split plot- CRD) design, e) a split plot - main plot in randomized complete blocks design (Split plot – RCB), f) a split-block or strip-split plot design, g) a split plot - main plot in row-column design (Split plot- 2 LS) (Pictures taken from Littell et al., (2008)).

Figure 1.a presents a **completely randomized design** (CDR) with a 2x2 factorial treatment structure. All the treatments are arranged in the field 4 times in a complete random way.

In Figure 1.b the experimental units have been arranged into r blocks. Each treatment is present in every block in a random way. This gives a **randomized complete block** **design** (RCB).

Figure 1.c presents a **latin square design** (LS) where the plots have been blocked on row and columns. In this design, each treatment should be present once in each row and each column.

### Split plot type experiments

In practice, some factors may be more difficult to vary than others at the level of experimental units. For example, typically irrigation and tillage can only be applied at a scale larger than the small plots. Then, it is more efficient to apply a difficult-to-change factor to the units at the top of the hierarchy and then apply the easier-to-change factor to a nested unit. This is called a split plot design. Figures 1d-1g show variations of split-plot designs. Generally, the 16 units of this 2×2 factorial experiment are grouped into 8 pairs of units, which appear as 1×2 rectangles, shown by the heavy dashed lines. The two levels of factor A (white vs grey) are randomly assigned to 4 of these larger units each. Within each pair, the two levels of B (lines vs no lines) are randomly assigned, one to each unit within the pair. This is a split-plot experiment. The units to which levels of A are assigned are called **whole-plot i.e. main plot- experimental units**, and the units to which levels of B are assigned are called **split-plot i.e. subplot-experimental units**. In general, a split-plot experiment is a factorial treatment design conducted such that the experimental unit with respect to one or more factors is a subunit of the experimental unit with respect to other factors.

The split plot shown in Figure 1.d uses a completely randomized design to assign levels of factor A to main plot experimental units. It is a **split plot- whole plot completely randomized (Split plot- CRD) design**

Figures 1.e, f, and g show variations of the split plot **using blocking.**

In Figure 1.e, the main plot experimental units are blocked as they were in Figure 1.b. The levels of A are assigned to units corresponding to half of the block (marked by the heavy dashed line), and the levels of B are assigned to the two units within each level of A. This is a **Split plot** experiment where **main plot is in randomized complete blocks. (Split plot – RCB)**

In Figure 1.f, the blocks consist of 2×2 squares. The levels of A are randomly assigned to the upper or lower half of each block, whereas the levels of B are randomly assigned to the right or left half. Depending on the author, textbooks may refer to this as a **split-block or strip-split plot.**

Figure 1.g uses a latin square to arrange the main plot experimental units. There are two 2×2 latin squares, one on the left (rows and columns 1 and 2) and one on the right (rows and columns 3 and 4). Levels of A are assigned within each square consistent with the constraints of the latin square. Each row-column combination has 2 split-plot units: levels of B are randomly assigned, one level per unit. This is a **Split plot** where main plot is **in row-column design (Split plot- 2 LS)**.

## **Structure of the excel file**

The excel file SoilCare\_experiments\_monitoring\_form\_vX consist of different sheets which include the different elements of the CS. For proper interpretation and similarity reasons many fields are predefined by the excel creator.

### Defining the relationships between each element

For practical reasons, compatibility with the structure of the database and in order to give the possibility for different relationships and combinations between the different elements of the experiment, in the sheets not only the information about each elements should be defined but also all the relationships among the elements.

For example one experiment may be conducted in different fields-locations, and at the same time in these fields, several experiments may take place. As well as one treatment of the experiment may be replicated in different fields/locations, and in one field treatments of various experiments may run. Because of these double relationships there is need to specify where each experiment is taking place and which field is used for each experiment and each treatment. Thus, in the sheet TREATMENT the information about the different treatments should be filled, which have a unique IDs (automatically generated by the system), and in the sheet TREATMENTS-links should define the relation of the treatments with the experiments and fields.



Figure 2: A field – experiment –treatment relationship. The treatments of the experiment one are “located” in two different fields and the treatments of the second only the second field.

These type of relationships may occur between several elements. For monitoring a system properly, it is essential that these relationships be defined.

### Drop down menu and predefined values

For minimizing the possibilities of typos, all the fields that are repeated several times like the unique identities of experiments, treatments etc. include a drop down menu. You have the option to type your value too. In case the value does not match the data validation restrictions (e.g. identity that does not exist) that have been defined you will get an error message.



Figure 3: Excel error message when the validation restrictions are not covered.

Furthermore, for several fields there are predefined options to select from. Please try to choose from the options given.

In the case your value/information **is not included in the predefined list**, you should type it in the LOOKUPTables sheet at the end of the relevant column in order to be able to track it and include it in the database.

### Fields colour explanation

|  |  |
| --- | --- |
| **Light Blue** | Definitions and units/ format of each property to be filled |
| **Grey** | Codes of each property  |
| **Light Orange** | Fields to be filled by the SS partners |
| **Orange** | Fields where value is automatically generated by the system |
| **Red**  | Explanations about the elements of each sheet |

### Cell values explanation

Please fill in all the cells in light orange colour. In the case something is not applicable or does not exist in your case please fill NA. If something is not measured, please fill NM. In any other case, a value should exist in each cell.

The full dates should be in the ISO-compliant yyyy-mm-dd format (To format dates in this format, select "Custom" format and type in "jjjj-mm-dd" under "Type:").

The coordinates should be given in the global reference system WGS84.

## Sheets details

### Explanation of the file’s sheets

* **General Info:** general instructions about the format and the way to fill each sector.
* **Institution-Person:** includes information about the institution and the persons that are involved in the experiments and the project.
* **Experiment Intro:** includes the general information about the experiments and the link of each experiment with the persons involved.
* **Fields:** details about the fields (location) where the experiments take place.
	+ **Fields-links:** Every field should be linked to at least one soil profile and at least one weather station.

**!** please be sure to provide the same weather station ID in this document with the one in the SoilCare\_weather\_monitoring\_form\_vX excel file which include the weather data (historical and future).

* **Soil profile:** includes information about the soil profiles and the properties /indicators that are derived from them, as described in the Report 10, D4.2 Monitoring Plan for the Study Sites. In the lower batch the detailed information and properties of each identified horizon should be given.

**!** For the soil profile, a picture should also be included

**!** For each field at least one soil profile should be assigned in order to provide a full characterization of the “baseline” information.

* **Treatments:** In this sheet, information about the different treatments of each experiment should be filled.
	+ **Treatments-links:** As mentioned each treatment may be located in different fields thus the information about each one may vary. Also in the same field different treatments of various experiments may run. Thus the triple connection experiment-field-treatment should be filled in the Treatments-links sheet.
* **Plot:** the layout of each plot (experimental unit) should be filled in this sheet as well as the links with the experiments and treatments. You should fill the information according to your design. E.g. if you don’t have a blocked design the block\_no should be NA or if you don’t have split plot design the main\_plot should be NA.

**!** The layout should also be included as an image in the format presented in the Picture format section below. An example of a layout is presented in Figure 4.



Figure 4: Experimental design example.(Presents a randomized complete block design (RCB) in which 12 treatments (center number) allocated in 48 experimental plots (plot number right corner) which are blocked in 4 blocks )

* **Factors:** In this sheet, the information about the primary factors (**those** **which change among the treatments and their effects are to be tested**) and their different levels are mentioned. Each treatment consists of a unique combination of different levels of the various factors. Here, the levels of each factor are described.
	+ **Factors-links:** The link of each level with the relevant treatment should be given.

After the sheet FACTORS-links, there are several sheets of management categories (e.g. Irrigation, fertilization, plantings etc.) and the sheets with their events (e.g. Irrigation-events, fertilization-events, plantings-events etc.). In these sheets, you should fill information about the historical and current **standard farm operations**, and the historical and current **treatments’ factors levels**.

**The standard farm operations** (management practices): procedures part of the cropping system which are not part of the experimental design.

**!** The treatments receive different management practices – farming operations whose effects are not tested. (E.g. The whole field may receive the same type and amount of fertilization and irrigation). Those “operations” are part of the cropping system but are not tested as part of the treatments. They **should only** be included in the relevant sheets (IRRIGATION, FERTILIZATION etc.) and in the sheets with their events.

**The treatments’ factors**: and the details of each different level of the factors and the events/applications should be included in the relevant sheets.

|  |
| --- |
| **! Information about the historical operations land use / crop history, or the levels of the experiment should be filled for at least the last 5 years (even without all the details-if you do not know) in order the history of the fields and experiments to be known.**  |

In order to fill the management sheets you should add the IDs on the levels (field, treatment, plot\_ where the changes are applied and refer to (factors or operations), and then at the lower levels you should write NA. (See example below).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CASE** | **field\_ID** | **treatment\_ID** | **plot\_no** | **factor\_level\_ID** | **operation\_ID** |
| **Factors’ levels (history & current)** | Field ID | Treatment ID | NA / plot no | Level ID | NA |
| **Operation (current & history)** | Field ID | Treatment ID / NA | NA / plot no | NA | Operation ID |

Table : Possible options to fill the management sheets

|  |
| --- |
| **Example**For example in a field there is an experiment running where each treatment receives different levels of organic material application, and every year also gets tilled (as a standard farm operation). Then you should fill in the ORGANIC\_MATERIALS sheet the detailed information about each organic material level and in the sheet TILLAGE the details about the tillage operations in the field. Then you should fill in the sheets ORGANIC\_MATERIALS -events and TILLAGE-events the details about each fertilization and tillage event, with their historical events in the previous 5 years. Thus, fertilization is a factor with levels, and tillage is and operation in the whole field. The relevant sheets you should fill are presented below.The sheet FACTORS includes the factor levels: Then in the sheet ORGANIC\_MATERIALS you should fill the details for each level. The operation\_ID should be filled as NA. Then the sheet ORGANIC\_MATERIALS-events should be filled with the information of each application of organic materials for each level. The lowest level on which the OM changes is per treatment thus plot\_no is filled with NA. In case the information change in one plot the information should be filled per plot. Tillage is a standard farm operation common in the whole field. Thus the sheet TILLAGE is filled like in the example:  You should give an operation\_ID. Then the lowest level on which tillage is changing is per field. Thus treatment\_ID should be filled as NA as well as factor\_level\_ID should be filled as NA as tillage is operation here. Then in the sheet TILLAGE-events you should fill the historical and current tillage events: Treatment\_ID, plot\_no and factor\_level\_ID should be filled with NA as tillage is an operation and the lowest level on which changes is per field. |

The sheets of management categories are presented below:

* **Plantings (events):** crop rotations, cover crops, planting or harvesting of crops should be filled in this sheet
* **Irrigation (events)**
* **Fertilizers (events)**
* **Organic materials (events):** application of various organic materials residues etc. should be filled in this sheet
* **Tillage (events)**
* **Chemicals (events)**
* **Mulches (events)**

After the management categories there are the following sheets:

* **Diary:** in the diary sheet, you should mention any important events that may take place, which can influence the results of the treatments and are not mentioned in any other element. For example here you can mention extreme weather events, damage in a plot by tractors, insects etc.
* **Observations-metadata:** The sheet includes the legend of the variables that are monitored. There are predefined variables and blank cells at the end where you can add variables for which you have data and not included in this predefined list. In this table, you should also mention the measurement method (or type of instrument) of each parameter used and remarks for the measurements or methods if any. If the sampling depth is different than the one mentioned in the Report 10, D4.2. then you should mention it in the remarks.
* **Observations-results:** The results of the monitoring, the values of each parameter should be filled. At the end of the predefined columns there are empty columns where your added in the list parameters will automatically appear. Please be sure to fill the values in the predefined unit. Please **do not** add the values of each replication within the experimental plot but the average value of the plot. Be careful with the use of Not Measured (NM) and Not Applicable (NA).

|  |
| --- |
| **!** Please add also observation results from previous years if you have (up to 5 years) even if different methods used in order the initial conditions and history to be known. |

* **Observation Notes:** Here a summary table of all the indicators to be measured/estimated/calculated is presented with details about the way of assessment/measured and their spatial resolution. You can consult this sheet in parallel with the Report 10, D4.2 Monitoring Plan for the Study Sites and the additional annexes received from WP4.
* **Documents:** Information about existing documents and publications about your experiments should be filled here.
* **Photos:** The metadata of the photos to be send should be filled here.
* **Lookup Tables:** The predefined list are presented here. Here you can also add in the corresponding table options that fit your case and are not already included.

### Details about specific indicators.

**Texture and Water Retention Characteristics (WRC) per horizon:** As mentioned in the Monitoring Plan final table of agreed information received by WP4, texture and water retention curve should be measured in each soil horizon of the soil profile. Thus the results of the measurements should be filled in the sheet SOIL\_PROFILE and then the relevant information in the OBSERVATION sheet should be marked as NM (not measured). In case you are not able to follow the protocol (which is not recommended) you should do the measurements in the top soil and add the result in the OBSERVATION sheet and in the relevant fields in the SOIL\_PROFILE sheet to mark as NM.

Furthermore, for the WRC if you have more points of the curve than the predefined in the Report 10, D4.2, please add these at the empty columns in the SOIL\_PROFILE sheet (if the measurements are per horizon) or as a new indicator in the OBSERVATION sheet (if the measurements become on the top soil).

**Yield quality indicators:** Yield quality can be characterized with different indicators according the crop. Please, add at least one yield quality parameter for each crop of your treatments with the relevant information in the remarks.

### Picture format

For the soil profile as well as the experimental design and plot arrangement pictures should be provided in a separate folder. Furthermore, you can provide other pictures with important context if you think that will contribute to the monitoring. All the photos should be in a folder together and send together with the filled excel file with the same unique name mentioned in the sheet PHOTOS. The photos should be in .jpg format and their metadata should be included in the sheet PHOTOS.

# References

1. Casler, M. D. (2015) ‘Fundamentals of experimental design: Guidelines for designing successful experiments’, *Agronomy Journal*, 107(2), pp. 692–705. doi: 10.2134/agronj2013.0114.
2. Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D.; Schabenberger, O. (2008) *SAS for Mixed Models*. Second. Cary, NC.: SAS Institute.
3. Oehlert, G. W. (2003) *A First Course in Design and Analysis of Experiments*, *The American Statistician*. doi: 10.1198/tas.2003.s210.
4. Parolini, G. (2015) ‘Charting the history of agricultural experiments’, *History and Philosophy of the Life Sciences*. Springer International Publishing, 37(3), pp. 231–241. doi: 10.1007/s40656-015-0079-5.
5. SAS Institute Inc. (2005) ‘Concepts of experimental design.’, *Biostatistics for animal …*, p. 34. Available at: <http://www.cabdirect.org/abstracts/20083014785.html>.

# Part 2: Document for the historical and future weather data

## File format

For each weather station used to characterize a field a different file should be used. Please be sure to give a unique weather station identity in each weather station and link those properly with the corresponding field. One weather station can be linked to several fields (if appropriate) and if possible, multiple weather stations can be linked to one field.

The file uses the same field colour format as the SoilCare\_experiments\_monitoring\_form\_vX file (explained in Part 1 of this document). The WEATHER sheet includes a metadata table where the required information of the weather station should be filled. The second table includes the legend of the variables that are monitored. There are several predefined variables and blank cells at the end where you can add variables for which you have data and not included in this predefined list. In this table, you should also mention the measurement method (or type of instrument) of each parameter from the weather station used and remarks for the measurements if any.

In the sheet WEATHER-daily, the daily values of each parameter should be filled. At the end of the predefined columns there are empty columns with a drop down menu to add your additional variables. Please be sure to fill the values in the predefined unit.

The same file format will be used to monitor the **historical daily data** to be sent (if possible up to 30 years) and afterwards the annual daily weather data of each station.

## Selection of the meteo-stations

For the study-sites, especially under dryer conditions, it is important to have a minimum set of nearby weather-data.

It is useful to have longer term data ( e.g. 30 years), characteristic for the climatic conditions, and data during the experiments from a nearby weather station.

The nearby weather data could be collected in a longstanding meteo-station, which is part of the official national meteorological measurements. Some experimental stations have a meteo-station, which is part of the national agro-meteorological network. However, especially for on-farm experiments this might not be the case.

For a large number of stations daily data are available on the European Climate Assessment & Dataset (ECAD) project: <https://www.ecad.eu//dailydata/index.php>. There is a specific search in this database possible based on latitude/longitude, station name. possible on: <http://climexp.knmi.nl/selectdailyseries.cgi>. WP5 can help to identify the nearest station in the ECAD.

For every SS it is useful to check which long-term station is representative for the SS. With the coordinates of the SS the most nearby station can be found.

If there is no such long-term station close to the SS then other shorter term series are needed. This is especially important for rainfall, as local heavy rain might have an impact on the experiments. The shorter nearby term series can be compared to the longer term representative series. If there is no station active in the immediate vicinity of the experiment(s) one could check on the nearby private weather stations (PWS) + airport weather-stations on: <https://www.wunderground.com/>. However, be very careful as PWS might not be delivering high quality data. It is advised to contact the owner of the PWS and visit the station.

## Needed meteorological data

Needless to say that daily rainfall is an important variable for the soil water balance.

Ideally daily data are available which are needed to the estimation of the crop evapotranspiration. The calculation by Penman-Monteith (recommended by FAO) we need daily values for:

Daily Tmax, Tmin, Humidity, Wind speed, Solar radiation.

For the formula by Hargreaves-Samani only daily Tmin and Tmax is needed. Only providing Tav is not so meaningful. Tmax and Tmin in a day contains a lot more information. In fact, from Tmin and Tmax we can estimate the humidity, the evapotranspiration, the solar radiation.

Please note that relative humidity (often expressed as percentage) is highly variable during the day. If only RH% is available the average RH% is not meaningful. Therefore it is often preferred to supply dew point. If only RH% is published, it is important to supply both RH% max and RH% min. In a lot of (humid) climates the RH% max is close to 100% and is observed at sunrise. In that case the dew point can be approximated by the Tmin. For older manual weather stations the instruments have been monitored two times at a specific hour in the morning and in the afternoon. Humidity can be cross-checked with Tmax and Tmin.

For solar radiation and wind-speed one has to be particularly careful. Both sensors can deteriorate and measure too low quantities. In addition wind-speed can be measured in a sheltered location. Most stations will measure at 10m. So in addition to the type of wind-speed sensor also the height of the sensor is crucial.

Also with automatic stations solar radiation can be measured by a large range of possible sensors capturing different wavelengths with different efficiencies. So information on the sensor, units for solar radiation and the age of the sensor is needed. Solar radiation can also be cross-checked with the Hargreaves-Samani estimate for solar radiation.

**!! So, the absolute minimum set is daily Rainfall, Tmax and Tmin. In addition if available, data on Humidity, Wind speed and Solar radiation are very needed.**

# Contact- questions/remarks

Please feel free to mail: ioanna.panagea@kuleuven.be, or Skype "ioannapanagea" or call: +32 16 37 67 83 if you have questions, comments and suggestions.