Carboseq WP2 Export Module Instructions

2022-07-12

Authors: Guillaume Blanchy, Ioanna Panagea, Greet Ruysschaert, (EV-ILVO)

1. Introduction

The export module is composed of multiple tabs. After loading the database using the 'Upload' button or copying the Google Sheet URL, a map with the locations of the data collected will be plotted (Figure 1). The following tabs will automatically appear:

- **Filtering**: filters the entire dataset according to user-defined criteria. The filtered dataset will be then used in the following tabs;
- **Query**: helps to define the 'control' and the 'treatment' factors under comparison, i.e., the *pair* hereinafter;
- **Export**: enables to export the selections made in the **Filtering** or **Query** tabs in 2 different formats (1. Filtered DB" format which is an .xlsx file with similar tab structure as the database and / or 2. a "Flat DB" format which is an .xlsx file with one single tab in which the information is stacked)
- **Selection**: enables exporting a selection of predefined experiments and treatments uploaded as an excel file with two columns ('Experiment ID' and 'Treatment ID') or three columns ('Experiment ID', 'Treatment ID_C', 'Treatment ID_T'). The database will be filtered accordingly and the data will be available in 2 different formats (1. Filtered DB" format which is an .xlsx file with similar tab structure as the database and / or 2. a "Flat DB" format which is an .xlsx file with one single tab in which the information is stacked).

There are two options to export data:

1. by **Filtering and/or Querying** and then exporting the selection. This option should be used to explore the database and find the possible relevant entries for your task.

2. by using the **Selection** tool with the predefined selection of experiments and treatments. This option should be used when you have selected all the relevant experiments and treatments in order to export the complete dataset to be used in the analysis.

When you are ready to export the dataset to be used in your analysis you should inform us. We will create a version of the database at that moment to be stored in a repository with a unique DOI number.

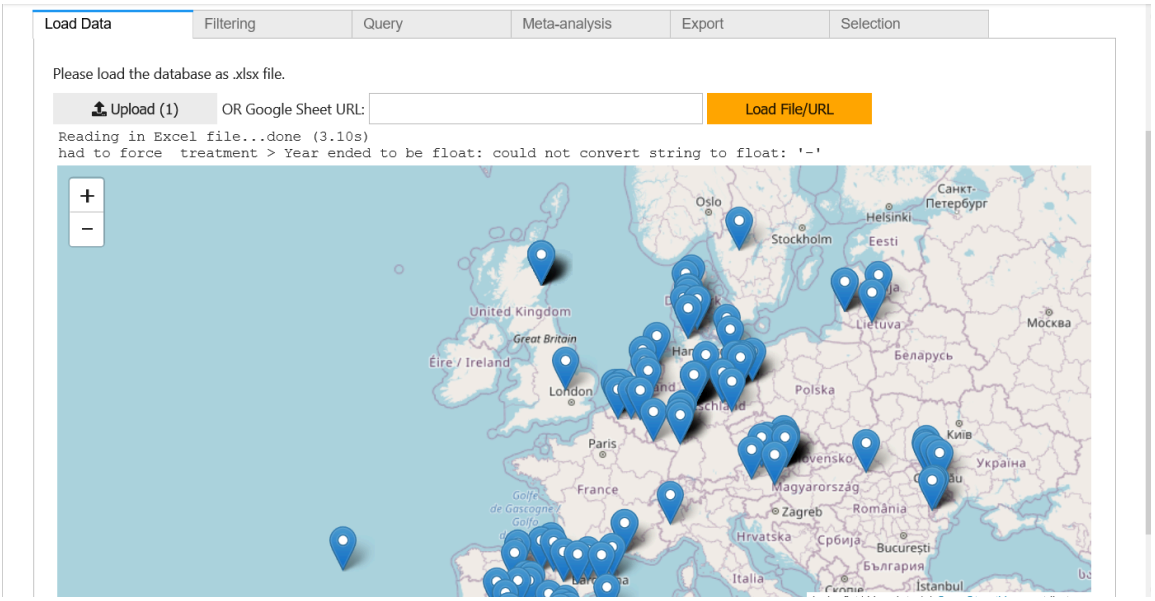

Figure 1: Screenshot of the interface of Carboseq WP2 Export Module after loading the input database.

2. Filtering the database

In this tab you can add filters to the data. The filters are always applied on the initial database uploaded. You can then further explore the filtered data by creating queries (Query tab).

For example you can select only the experiments which are located in a specific country (here Austria) and then proceed to query only those in the Query tab.

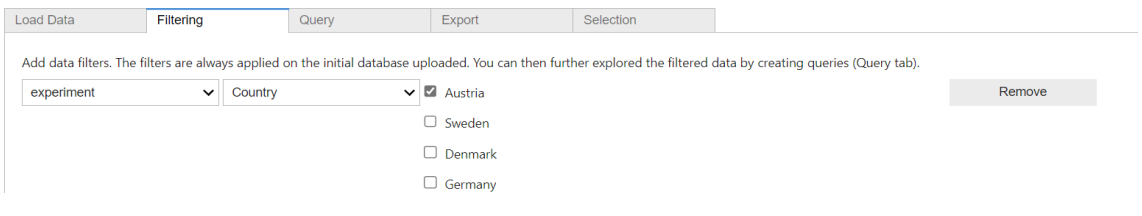

Figure 2: Filtering from the initial database only the experiments which are located in Austria.

You can apply several filters at the same time. For example you can filter the initial database and select only the experiments that belong to a specific climatic zone and have a specific practice. Eg Zero tillage in the Mediterranean climatic zone.

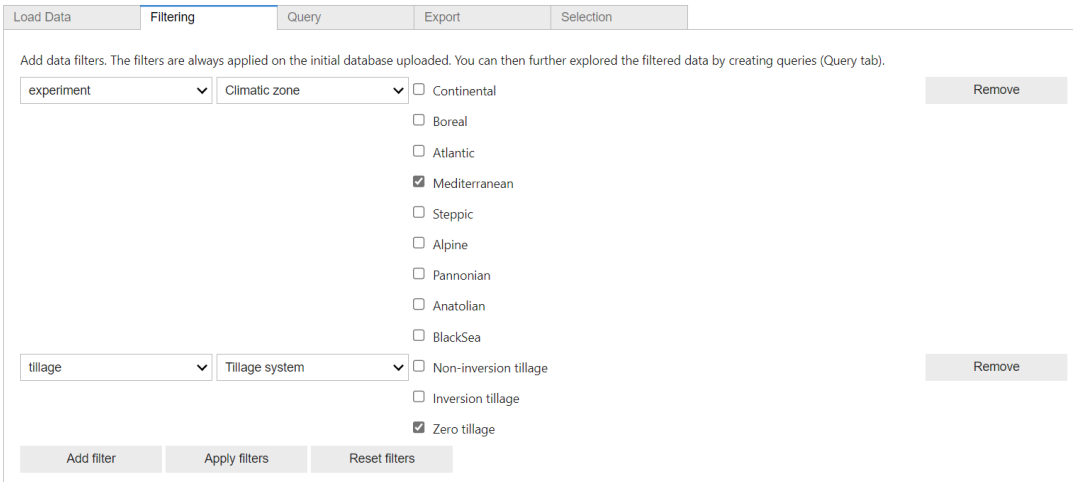

Figure 3. Filtering the database to export or query those experiments that belong to a mediterranean climatic zone and include zero tillage system.

3. Querying the database

The database does not contain any explicit and structured information about the factors to be investigated. Great attention was paid to grant the flexibility to define specific research questions to the user, who can query all experiments potentially answering to his/her research questions.

Any query is defined as a **pair**, constituted by a ***control*** and by a ***treatment***); is formed by a reference/control (button 'Add reference') where you define the sheet, column and value you want. You can then specify the '**occurrence**' of the value (this value against other, presence or absence, 1 vs more occurence of the value).

The query algorithm searches for every pair of treatments matching the pair definition. It extracts the reference/control rows and looks for their corresponding 'treatment' rows, thus forming pairs of 'control vs treatment'.

To further restrain the comparison, **conditions** can be added to force both 'control' and 'treatment' rows to share the same or a specific value for a given column.

For instance, let us consider an experiment where four treatments were tested, i.e., ‘cover crop + zero tillage’, ‘cover crop + inversion tillage’, ‘no cover crop + zero tillage’, ‘no cover crop + inversion tillage’. The pair defined in the query include the two terms ‘zero tillage’ vs not ‘zero tillage’ (= ‘inversion tillage’ in this case) as comparison. The query algorithm will find the following pairs:

- ‘no cover crop + zero tillage’ (control) vs ‘no cover crop + inversion tillage’ (treatment)
- ‘cover crop + zero tillage’ (control) vs ‘cover crop + inversion tillage’ (treatment)

3.1 Forming a query and selecting the proper occurrence

The query is formed by specifying the **control value**. For instance, if we want to compare ‘Inversion tillage’ with ‘Non-inversion tillage’ (or ‘Zero tillage’), ‘Inversion tillage’ must be chosen as the control value.

Then, the type of comparison is defined as one out of three types of **occurrence**:

- ‘**This value vs other**’: for instance, ‘Inversion tillage’ vs other tillage system (so ‘Zero tillage’ or ‘Non-inversion tillage’). If there is at **least one row** that contains a different value, the treatment will be considered eligible in the query.

Table 1: Example of ‘this value vs other’ for ‘tillage’ sheet with one row per treatment.

Experiment ID	**Treatment ID**	**Tillage system**	
cctillage	treatment1	Inversion tillage	-> control
cctillage	treatment2	Zero tillage	-> treatment

Table 2: Example of ‘this value vs other’ for ‘tillage’ sheet with multiple rows per treatment.

Experiment ID	**Treatment ID**	**Rotation**	**Tillage system**	
cctillage	treatment1	Year 1	Inversion tillage	-> control
cctillage	treatment1	Year 2	Inversion tillage	
cctillage	treatment2	Year 1	Inversion tillage	-> treatment
cctillage	treatment2	Year 2	Zero tillage	

=> Note: in this case, treatment2 might be wrongly classified as the ‘treatment’ because it is not every year ‘Zero tillage’. => This example illustrates that the extraction of the relevant experiments by the export module must be always checked by the user

Table 3: Even when multiple tillage options are specified, if at least one of them is different from the control, then it is considered as treatment.

Experiment ID	**Treatment ID**	**Rotation**	**Tillage system**	
cctillage	treatment1	Year 1	Inversion tillage	-> control
cctillage	treatment2	Year 1	Inversion tillage	-> treatment
cctillage	treatment2	Year 1	Zero tillage	

- ‘**Absence vs presence**’: for instance, the absence of cover crops (set as control) vs the presence of cover crops (inside the rotation year, or not)

Table 4: Example of ‘Absence vs presence’ for ‘crops’ sheet with an unequal number of rows for the treatments.

Experiment ID	Treatment ID	Rotation	Crop type	
cctillage	treatment1	Year 1	Main crop	-> control
cctillage	treatment1	Year 2	Main crop	
cctillage	treatment2	Year 1	Main crop	-> treatment
cctillage	treatment2	Year 2	Cover crop	
cctillage	treatment2	Year 2	Main crop	

Table 5: Counter example of ‘Absence vs presence’ for ‘crops’ sheet. Here no control can be defined as **all** the rows of the control need not have the value (‘Cover crop’ in this case) **for all rotation years**.

Experiment ID	Treatment ID	Rotation	Crop type	
cctillage	treatment1	Year 1	Main crop	-> not a control as some rows contain the value ‘Cover crop’ that we want to test its absence
cctillage	treatment1	Year 2	Main crop	
cctillage	treament1	Year 2	**Cover crop**	
cctillage	treatment2	Year 1	Main crop	
cctillage	treatment2	Year 1	Cover crop	
cctillage	treatment2	Year 2	Main crop	
cctillage	treatment2	Year 2	Cover crop	

- ‘**1 occurrence vs more**’: for example: one main crop or multiple main crops **within the same rotation year**. If not set, the condition that the ‘Rotation’ columns must be identical between the elements of the pairs is automatically set.

Table 6: Example of ‘1 occurrence vs more’ for ‘crops’ sheet. In treatment2 there are two ‘Main crop’ for ‘Rotation’ year 2.

Experiment ID	Treatment ID	Rotation	Crop type	
cctillage	treatment1	Year 1	Main crop	-> control
cctillage	treatment1	Year 2	Main crop	
cctillage	treatment2	Year 1	Main crop	-> treatment
cctillage	treatment2	Year 2	Main crop	

cctillage	treatment2	Year 2	Main crop	

3.2 Adding specific conditions to the query

In addition to the control value, it is also possible to add **conditions**. Conditions can be of two types:

- we force the two treatments of the pairs to have the **same value** for one column (e.g. the same tillage system).

A **special case** is when we force the **'Rotation'** to be the same, this means that we will only look for pairs **inside every rotation year**. If a pair is formed in at least one rotation year, then the experiment is selected.

Table 7: 'Absence vs presence' of cover crops with the condition that the 'Cropping system' must be the same for each element of the pair inside the same rotation year. There are two conditions here: (1) the cropping system has to be the same between control and treatment, (2) the rotation year ('Rotation' column) should be the same between control and treatment.

Experiment ID	**Treatment ID**	**Cropping system**	**Rotation**	**Crop type**	
cctillage	treatment1	Intercropping	Year 1	Main crop	-> control
cctillage	treatment1	Monoculture	Year 2	Main crop	
cctillage	treatment2	> Mixture intercropping	Year 1	Main crop	-> treatment (because 'Rotation' is forced to be the same and the conditions are met in Year 2)
cctillage	treatment2	Monoculture	Year 2	Cover crop	
cctillage	treatment2	Monoculture	Year 2	Main crop	
cctillage	treatment3	Monoculture	Year 1	Main crop	-> not selected as different 'Cropping system' from control
cctillage	treatment3	Intercropping	Year 2	Cover crop	
cctillage	treatment3	Intercropping	Year 2	Main crop	

- we force them to have the **same specific value** (e.g. they should both be 'Zero tillage'). This is implemented by first filtering all the rows that satisfy this condition in the sheet of interest.

A detailed explanation about the query algorithm can be found in the appendix.

3.3 Cases

This section presents a few examples on how to define a ‘control’ and a ‘treatment’ with the **‘Query’** tab.

3.3.1 Inversion tillage vs other tillage type

At first a control value (also called ‘reference’ value) is defined. In this example, the control value is in the sheet ‘tillage’, in the column ‘Tillage system’ with the value ‘Inversion tillage’. Then the type of comparison (type of ‘occurence’) can be chosen. In this case, the occurrence is ‘This value vs other’. Which means that for a given Experiment ID, the treatments with ‘Inversion tillage’ will be compared to other treatments that **do not have** ‘Inversion tillage’. In addition, conditions can be added that force the two elements of the comparison pair to have some columns or value in common. In this example, both elements of the pair **must have the same** ‘Land use’.

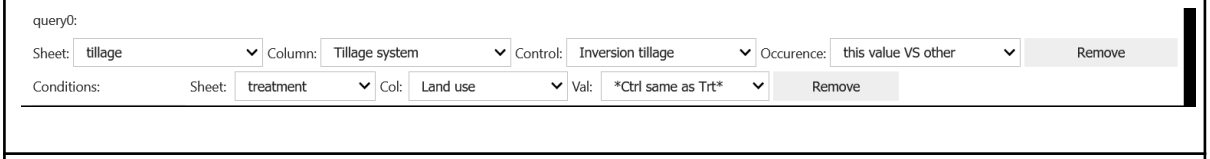

Figure 4: Example of inversion tillage vs not inversion tillage (so zero tillage and non-inversion tillage) within the same land use.

3.3.2 Cover crops vs no cover crops

In this example, the **absence** of cover crops (control) is compared against the **presence** of cover crops (treatment). ‘Tillage system’ should **be the same for both** and the ‘Cropping system’ must be ‘Monoculture’ for both.

The query algorithm will look at all the rows in the ‘crops’ tab where the ‘Cropping system’ is ‘Monoculture’ AND in the ‘tillage’ tab where the ‘Tillage system’ is equal (option ‘Ctrl same as Trt’). In this case, the occurrence is ‘Absence vs presence’.Inside this subset of rows, it will look by Treatment ID for one that **does not have any** ‘Crop type’ equals ‘Cover crop’ (control is the absence of the value) vs other Treatment ID that have at least one ‘Crop type’ equals ‘Cover crop’. This is independent of the ‘Rotation’ year.

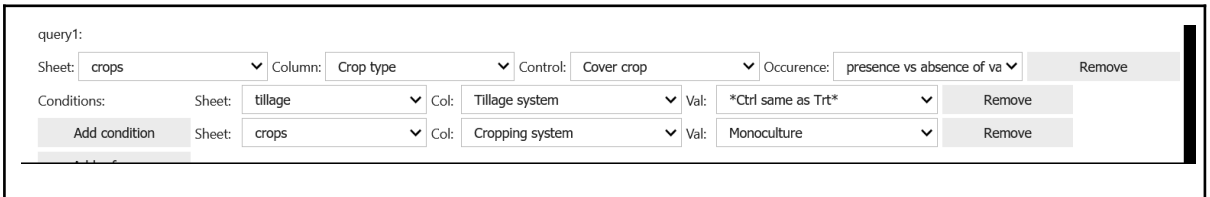

Figure 5: Absence vs presence of cover crops with the same tillage system and for monoculture only.

3.3.3. One crop vs multi-cropping (several crops per year)

This query identifies pairs containing **one** crop in a rotation year **vs multiple** crops within the same rotation year.

The query algorithm will first find at least one rotation year that contains exactly one occurrence of the specified crop. It will then look at other treatments and will try to find the treatment that for the same rotation year contains more than one occurrence of the specified crop.

Figure 6: One crop vs multiple crops within the same rotation year.

3.3.4 Cover crop x vs other cover crops

This query aims at finding treatments in which another cover crop than cover crop X is present.

The query algorithm will first filter all rows of the ‘crops’ sheet to only keep the one where ‘Crop type’ equals ‘Cover crop’. Then it will look for a control Treatment ID where ‘Crop’ equals ‘> Lolium/Ryegrass’ and then a corresponding Treatment ID where ‘Crop’ is not equal to ‘> Lolium/Ryegrass’ to form the pair. All comparisons are done within the same rotation year.

Figure 7: Cover crop ‘Lolium’ vs other cover crops within the same rotation year.

3.3.5 Different termination/harvesting methods

The query algorithm will first keep only the rows of the ‘crops’ tab where ‘Crop type’ is ‘Cover crop’. Then it will look, within a rotation year, for rows with a ‘Harvesting/Termination method’ equals to ‘Mowing’ and try to find other rows from another treatment that have not ‘Mowing’ as ‘Harvesting/Termination method’ to form a pair.

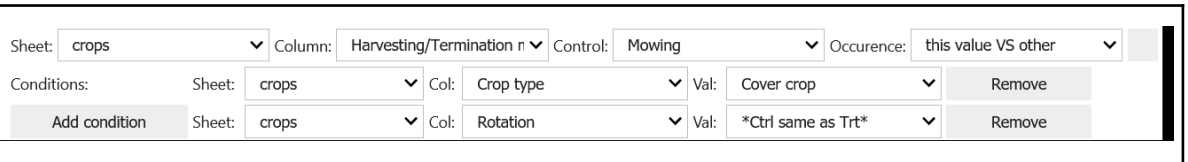

Figure 8: Different termination strategy for cover crops.

3.3.6 Different harvesting period

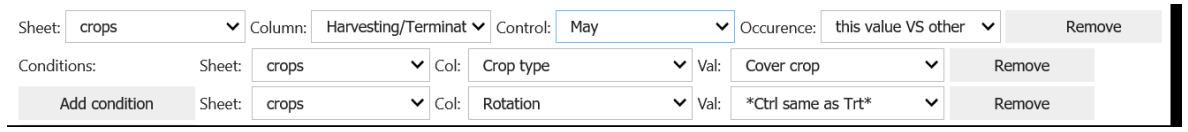

Figure 9: Different harvesting/termination period for cover crops.

3.3.7 ‘Full residue removal’ vs ‘No or Partial residue removal’

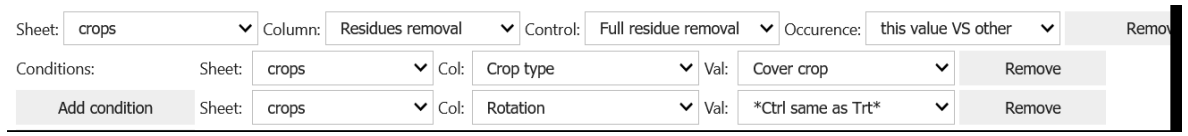

Figure 10: Full residue removal vs ‘no residue removal’ or ‘partial residue removal’ within the same rotation year only for cover crops.

3.3.8 ‘Residues incorporated’ vs ‘Residues on surface’

Figure 11: Residue incorporation vs residues left on the surface for cover crops within the same rotation year.

4. Dealing with levels in Querying / Filtering tabs

Several drop-down menus in the template contain multiple levels such as in the ‘Type of fertilizer/amendment’ column of the ‘amendment’ sheet:

- Organic amendment
  - > Manure
    - >> Chicken manure
    - >> Park manure
  - > Compost
    - >> Farmyard compost
    - >> Woodchip compost
- Mineral amendment
  - > P-fertilizer

- > N-fertilizer
- > NPK-fertilizer

To enable comparison between higher levels, each column containing more than one level is splitted into different columns. For instance ‘Type of fertilizer/amendment’ will be splitted into ‘Type of fertilizer/amendment (l0)’, ‘Type of fertilizer/amendment, (l1)’, ‘Type of fertilizer/amendment (l2)’. Table 8 shows an example of corresponding levels. This way, it is possible to compare specific type of manure (e.g. ‘Pork manure’ vs ‘Chicken manure’ at level 2) or a more general group such as ‘Manure’ vs ‘Compost’ (level 1) or even ‘Organic amendment’ vs ‘Mineral fertilizer’ (level 0). Note that if a value is entered in the database at level 0 (e.g. ‘Organic amendment’), it will be copied to deeper levels as well. This is to enable comparison of the type ‘this value vs other’.

Table 8: Correspondence of value entered and equivalence in the different levels columns. Note that if higher levels are entered (e.g. ‘Organic amendment’ level 0), these are propagated towards all deepest levels (shown in italic).

Value entered	**Level**	**Level 0**	**Level 1**	**Level 2**
Organic amendment	0	Organic amendment	*Organic amendment*	*Organic amendment*
> Manure	1	Organic amendment	> Manure	> *Manure*
>> Chicken manure	2	Organic amendment	> Manure	>> Chicken manure
>> Pork manure	2	Organic amendment	> Manure	>> Pork manure
> Compost	1	Organic amendment	> Compost	> *Compost*
>> Farmyard compost	2	Organic amendment	> Compost	>> Farmyard compost
>> Woodchip compost	2	Organic amendment	> Compost	>> Woodchip compost
Mineral amendment	0	Mineral amendment	*Mineral amendment*	*Mineral amendment*
> P-fertilizer	1	Mineral amendment	> P-fertilizer	> *P-fertilizer*
> N-fertilizer	1	Mineral amendment	> N-fertilizer	> *N-fertilizer*
> NPK-fertilizer	1	Mineral amendment	> NPK-fertilizer	> *NPK-fertilizer*

5. Export Tab

This tab is used to generate and export the rows relevant to the queries or the entire database after filtering. Two export formats are available:

or a "Flat DB" format which is an .xlsx file with one single tab in which the information is stacked. If queries were made, the "Flat DB" format will contain columns with _C (for control) and _T for treatment so that this file can be used for data analysis more easily.

- Filtered DB: this format is similar to the template (a .xlsx with multiple tabs) and just keep all rows relevant to the queries made or the filters applied (that contains the Experiment ID and Treatment ID)
- Flat DB: this format is an .xlsx file with one single tab in which the information is stacked. If queries were made, the "Flat DB" format will contain columns with _C (for control) and _T for treatment so that this file can be used for data analysis more easily.

- Note that:
    - if data are available from multiple 'Sampling year', the rows will just be stacked (one row per sampling year) and other metadata will be repeated. This way, we let the user decide which data from which year to use for the meta-analysis.
    - If several metadata (= multiple rows) are available for one Treatment ID (e.g. multiple crops), they will be concatenated with " | " (e.g." >> Winter wheat | > Potatoes | > Clover"). If all the rows have the same value, only this value is kept once (e.g. ">> Winter wheat | >> Winter wheat | >> Winter wheat" becomes just ">> Winter wheat").

6. Selection Tab

In addition, it is possible to have a selected list of Experiment ID and Treatment ID and use those to filter the database ("Selection" tab of the export module). In this case an excel file (.xlsx) with two ('Experiment ID', 'Treatment ID') or three ('Experiment ID', 'Treatment ID_C', 'Treatment ID_T') columns is supplied. Two outputs are available: a filtered DB and a flat DB format which contains both '_C' (control) and '_T' (treatment) columns if a three column .xlsx was uploaded for the selection.

If you want to export a flat DB with a selection of columns, you can go first in the Export tab, make your selection of columns to be exported and then in the Selection tab to upload your excel file with the predefined experiments and treatments list. Then the flat DB format will only include the selected columns and the columns from the "data-soil" and "data-crop" tabs and also columns relevant to the "Data entry person" which are included by default.

	A	B
1	**Experiment ID**	**Treatment ID**
2	Melle2	PG
3	Melle2	TG
4	Melle2	TA
5	Melle2	PA
6	KUNZO13V59.1	Field B+NPK
7	KUNZO13V59.1	Field B +FYM
8	KUNZO13V59.1	Field B +FYM+NPK
9	KUNZO13V59.1	Field B +Csl+ST
10	KUNZO13V59.1	Field B +Csl+ST+NPK
11	KUNZO13V59.1	Field IV +NPK
12	KUNZO13V59.1	Field IV +FYM
13	KUNZO13V59.1	Field IV +FYM+NPK
14	KUNZO13V59.1	Field IV +Csl+ST
15	KUNZO13V59.1	Field IV +Csl+ST+NPK

Figure 12a. Predefined selection using only Experiment ID and Treatment ID

	A	B	C
1	**Experiment ID**	**Treatment ID_C**	**Treatment ID_T**
2	Melle2	PA	PG
3	Melle2	PA	TG
4	Melle2	PA	TA
5	KUNZO13V59.1	Field B+NPK	Field IV +NPK
6	KUNZO13V59.1	Field B +FYM	Field IV +FYM
7	KUNZO13V59.1	Field B +FYM+NPK	Field IV +FYM+NPK
8	KUNZO13V59.1	Field B +Csl+ST	Field IV +Csl+ST
9	KUNZO13V59.1	Field B +Csl+ST+NPK	Field IV +Csl+ST+NPK

Figure 12b. Predefined selection using Experiment ID ,Treatment ID_C and Treatment ID_T to export ready for analysis pairs.

Appendix

Figure A1 summarizes the working of the query algorithm. Further details are available in the verbose explanation below.

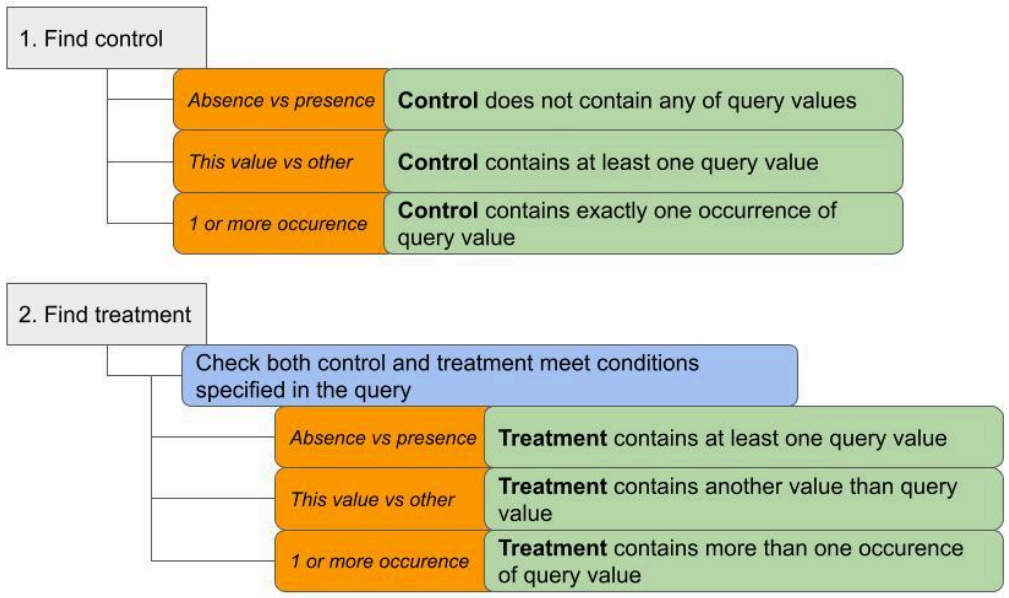

Figure A1: Simplified working of the query algorithm of the export module. Note that according to the occurrence value set ('absence vs presence', 'this value vs other', '1 or more occurrence'), different criterias will be used by the algorithm to find a suitable pair of control and treatment.

For each Experiment ID:

1. Find a Treatment ID that can be considered as a control
    a. Absence vs presence: control **does not contain any** of the query value (independent of the rotation year)
    b. This value vs other: control **contains at least one** query value (independent of the rotation year)
    c. 1 or more occurrence: control **contains exactly once** occurrence of the query value (**within the same rotation year**)
2. For each other Treatment ID (different from the control)
    a. Check if the control and the Treatment ID selected met the conditions in the other sheets. If the condition is 'same values' then this is first checked (e.g. if they have the same 'Tillage system') OR if the condition is the 'same specific value' then the rows with these values are first filtered before the occurrence is checked.
    b. Check the occurrence:
        i. Absence vs presence: check if at least one row of the selected Treatment ID **has the query value** (check per rotation year if we force 'Rotation' to be the same)

ii. This value vs other: check if at least one row of the selected Treatment ID contains **another value than the control** (check per rotation year if we force ‘Rotation’ to be the same)
iii. 1 or more occurrence: check if at least one row of the selected Treatment ID contains **more than one occurence** of the query value per rotation year

Note that the query algorithm will check for **at least one row** within the selected treatment (at least one row within the rotation year if ‘Rotation’ is enforced to be the same). This means that not all rows (within a rotation year or not) need to match the query for the two treatments to be selected. This will certainly produce irrelevant pairs but will ensure that all relevant pairs are in the selection.